



HAL
open science

Evolutionary genetics evidence of an essential, non-redundant role of the IFN- γ pathway in protective immunity

Jeremy Manry, Guillaume Laval, Etienne Patin, Simona Fornarino, Christiane Bouchier, Magali Tichit, Luis Barreiro, Lluís Quintana-Murci

► To cite this version:

Jeremy Manry, Guillaume Laval, Etienne Patin, Simona Fornarino, Christiane Bouchier, et al.. Evolutionary genetics evidence of an essential, non-redundant role of the IFN- γ pathway in protective immunity. *Human Mutation*, 2011, 32 (6), pp.633-42. 10.1002/humu.21484 . hal-00627541

HAL Id: hal-00627541

<https://hal.science/hal-00627541>

Submitted on 29 Sep 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Evolutionary genetics evidence of an essential, non-redundant role of the IFN- γ pathway in protective immunity

| | |
|-------------------------------|--|
| Journal: | <i>Human Mutation</i> |
| Manuscript ID: | humu-2010-0462.R1 |
| Wiley - Manuscript type: | Research Article |
| Date Submitted by the Author: | 23-Jan-2011 |
| Complete List of Authors: | Manry, Jeremy; Institut Pasteur Laval, Guillaume; Institut Pasteur Patin, Etienne; Institut Pasteur Fornarino, Simona; Institut Pasteur Bouchier, Christiane; Institut Pasteur Tichit, Magali; Institut Pasteur Barreiro, Luis; University of Chicago Quintana-Murci, Lluís; Institut Pasteur, EEMI |
| Key Words: | IFNG, IFNGR1, IFNGR2, natural selection, polymorphism, population genetics |
| | |

SCHOLARONE™
Manuscripts

RESPONSES TO REVIEWERS:**Referee: 1****Comments to the Author**

Manry et al examine genetics of the IFN-g pathway at a population genetics level. This study presents several interesting observations about selective pressures on the IFN-g pathway. The statistical methods are thorough and the resequenced dataset is valuable. Some weaknesses include an unusual selection of samples that may have a sample size and composition that is not ideally matched for the analytic goals of the project. In addition, publicly available Phase III HapMap data is not utilized effectively to enrich the data set and analysis.

Major Points

1. Populations and sample size: 186 individuals were resequenced for 3 genes. The 186 are selected from at least 11 populations. How were these numbers chosen and why were these populations chosen? The rationale is important and not articulated. It is important because the conclusions regarding purifying selection may have more or less generalizability based on the input populations. To make arguments about selection, it is important to have larger numbers of samples from distinct populations where evidence of selection can be observed. To characterize global SNP diversity and genetic effects of population migration, smaller numbers of samples from multiple populations is beneficial. Given that one of the primary goals of this project is to examine natural selection, the sample size for each population is very small (6 Orcadians, 4 Cambodians, 10 Japanese in particular). In particular, some of the selected populations are almost certainly composed of highly heterogeneous subpopulations. For example, the 33 Chinese minorities. Are these from geographically proximal or distant locations?

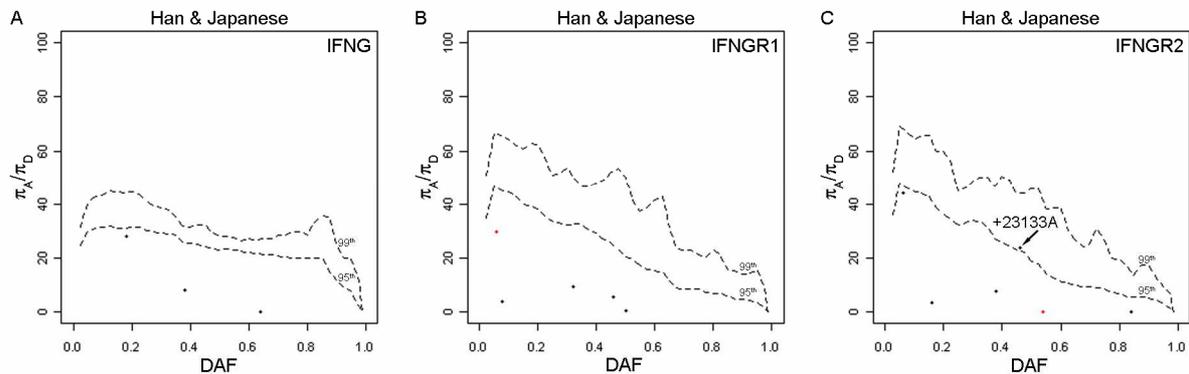
We thank the reviewer for this comment. With respect to the study design and sample choice, this was based on the double and complementary goals of our study. Indeed, as the reviewer suggests, one aim of our study was the discovery of new polymorphisms, in the line of one of the Human Mutation interests. To this end, we included individuals from many distinct ethnic origins. The second aim was to identify signatures of natural selection, where many individuals per continental region are needed. Our population choice reflects the need to balance between the two goals of the study. We think that this choice is the best compromise to allow us to address these two questions: several sub-populations to detect new polymorphisms, 186 individuals taken as a whole to detect signatures of natural selection species wide, and 62 individuals per continental group to detect more subtle signatures of local selection.

The reviewer is also right in wondering the extent to which this population choice may affect our results concerning natural selection. In particular, grouping and analysing together populations that present substantial genetic differentiation could generate biases in neutrality statistics. We have now formally tested this possibility by quantifying the levels of population differentiation among populations we grouped in our analyses, namely those living in the same continental region. Specifically, we have performed an AMOVA to estimate the fraction of the genetic variance of our dataset explained by (i) differences between individuals within a given population, (ii) differences among populations within the same continental region, and (iii) differences among the three continental groups (i.e. each group representing the merge of

1
2
3 the different populations of a given continent). Our analyses show that the fraction of the
4 genetic variance explained equals to 89.27%, 0.45% and 10.28%, respectively. In this view,
5 we can say that the genetic differentiation among populations from the same continental
6 region is negligible (mean=0.45%; African=0.73%, European=0.08%, and Asian=0.55%) and
7 non-significant in our dataset. In addition, and consistent with our data, genome-wide
8 genotyping datasets performed on the same individuals and populations here studied have
9 recently shown that the levels of population structure within continental regions is limited (Li
10 et al. 2008, Science). This is also true for other genome-wide datasets on similar populations,
11 e.g. the HapMap samples of Han Chinese and Japanese have been merged in all analyses due
12 to their high genetic resemblance (Frazer et al. 2007, Nature). Altogether, our analyses,
13 fuelled by the results of recent genome-wide datasets, indicate that the genetic differentiation
14 observed among subpopulations from the same continent is weak enough not to influence any
15 of our conclusions regarding natural selection acting in each continental population group.
16
17
18
19

20 With respect to the detection of purifying selection, the test used (MKPRF) is based on the
21 whole sample of 186 individuals merged together, because our aim was to detect how intense
22 have been the selective constraints (i.e. strong purifying selection against amino-acid changes,
23 estimated using ω) at the species-wide level (in all humans). Apart from that, this test is
24 insensitive to the number of populations sampled, because ω relies on the comparison of
25 amino-acid altering and silent sites, where silent sites represent an internal control of what is
26 expected in the absence of selection. Consequently, the sample design, whatever it is, will
27 equally influence both types of segregating sites and will not influence the parameter
28 estimation. In addition, the intensity of purifying selection on *IFNG* estimated from our
29 dataset ($\omega = 0.0189$) is extremely similar to that obtained using another population panel ($\omega =$
30 0.0184 from Bustamante et al. 2005, Nature), indicating that the detection of strong purifying
31 selection is not sensitive to the population considered.
32
33
34

35 With respect to the detection of positive selection, the reviewer is right in that the potential
36 presence of population substructure in some of our continental populations may influence
37 inferences concerning the effects of local positive selection. Indeed, in contrast with tests that
38 considered the human species as a whole (see MKPRF above), some of the tests aiming to
39 detect local positive selection are influenced by the structure of the studied populations, even
40 if low. For example, tests based on the allele frequency spectrum (Tajima's *D*, Fu & Li's *D**
41 and *F**) are sensitive to the amount of singletons, a feature that can reflect both positive
42 selection and population substructure (Ptak and Przeworski, 2002, Trends Genet). However,
43 we never made any claim of local selection based only on tests that are sensitive to population
44 substructure. For example, as the reviewer points out, the presence of Chinese minorities in
45 the East-Asian sample could be a problem in this respect. However, our results for the
46 *IFNGR2* +23133A allele, which is the only case for which we claim selection in Asia, kept
47 being significant when removing Chinese minorities for the DIND test and the F_{ST} statistics
48 (see analyses here below). For this analysis, we used the 2 major Asian populations only: Han
49 Chinese and Japanese (similarly to what was previously done to detect Asian specific
50 signatures of recent positive selection using the HapMap dataset (Voight et al., 2006, PLoS
51 Biol).
52
53
54
55
56
57
58
59
60

DIND test excluding the Asian minorities:

We have now clarified our choice regarding the sampling design in Material and Methods (page 5, and a new Supp. Table S1 to describe our population panel), and we have cautiously discussed the possibility of population structure in the discussion (page 18).

2. HapMap data: Phase 3 of HapMap has full genome sequencing on 11 populations and hundreds of individuals. Phase 3 the HapMap dataset is more extensive than the current paper. This paper appears to utilize some data from Phase II of Hapmap” but even this effort does not fully utilize the data. This is a lost opportunity to not use the Phase III Hapmap data and to fully analyze it in conjunction with the authors primary data.

We entirely agree with the reviewer in that the use of genome-wide datasets (e.g. HapMap, 1000 genomes, Celera, etc) is very useful for comparative purposes. Specifically, the reviewer points out that HapMap Phase III has sequenced 10 regions of 100kb on 11 populations and about 1000 individuals. However, there are a number of reasons that prevented us to make full use of the HapMap Phase III dataset: (i) the 10 resequenced regions do not include our 3 genes; (ii) because the remaining data (and this is valid also for HapMap Phase II) is based on genotyping only, we cannot perform neutrality tests such as Tajima’s D , Fu & Li’s D^* and F^* and Fay & Wu’s H , which need full resequencing data; and (iii) HapMap data is in turn useful for tests based on long-range haplotype homozygosity, in which case we had to use the HapMap Phase II dataset because the SNP density is greater (~3.1 million genotyped SNPs) than that of HapMap Phase III (~1.4 million genotyped SNPs). In the context of our paper, we now used the HapMap Phase III data for allele frequency comparisons (see point 3), obviously for those SNPs that have been genotyped in this dataset.

To circumvent the limitations associated with the nature of the HapMap data (i.e. SNP genotyping which is subject to ascertainment biases), one has to use datasets that are based on full resequencing (e.g. 1000 genomes, Celera dataset, etc). The 1000 genomes project is in a pilot phase that includes 3 distinct approaches (Durbin et al, 2010, Nature). The first is a trio project where 2 trios have been whole-genome sequenced at high coverage, a sample size that is too low to provide with reliable comparisons in the context of our study. The second is an exon-targeted resequencing project where 900 genes have been sequenced at high coverage in 697 individuals. In this case, the number of exploitable genes is much lower than that of other public datasets (see Celera dataset discussed below), especially when it comes to make

1
2
3 comparisons among different functional classes as the reviewer suggests. The third project is
4 a low-coverage whole-genome sequencing in 179 individuals. Although the use of the third
5 project of 1000 genomes project is seducing, there are a number of reasons for which it is
6 hardly comparable with our data: (i) the sequence coverage varies among populations, which
7 complicates comparisons of neutrality statistics among populations (see Suppl Fig 2 in Durbin
8 et al., 2010 Nature), (ii) the variation in coverage along the genome strongly limits
9 comparisons between genes within the same population, (iii) the low coverage of most of this
10 dataset (in average 3.6x) limits the detection of low frequency variants, which in turn,
11 represent the substrate to detect and estimate the intensity of purifying selection (i.e. one aim
12 of our study). Indeed, it has been shown that, because non-synonymous variants are generally
13 found at low frequencies (<5%), this dataset has reduced power to discover variants in this
14 range, and therefore alter interpretation as to selection pressures (Durbin et al, 2010, Nature).
15 In this view, we decided to use the Celera dataset, which contains 11,624 genes that have been
16 fully resequenced (Bustamante et al., 2005, Nature), for the following reasons: (i) it is, by the
17 time being, the resequencing project providing the largest number of resequenced genes
18 (11,624); (ii) genes have been resequenced using standard PCR-based techniques, thereby
19 excluding any of the limitations introduced by next-generation sequencing (e.g. coverage
20 variation), and (iii) this dataset has been already specifically used for the genomewide
21 detection of the intensity of purifying selection (Bustamante et al., 2005, Nature), constituting
22 therefore a perfect comparative dataset in the context of our study.
23
24
25
26
27

28 In the revised version of the manuscript, following the reviewers' and editorial advices, we
29 now compare our data in the context of the genome-wide Celera dataset (see Material and
30 Methods page 7, Results pages 12-13 and Discussion page 19), where we integrate our results
31 in the more general context provided by these 11,624 genes. In addition, we have now added
32 in Supp. Table S3 the SNPs that are also present in HapMap Phase III. We have also
33 complemented this comparison with a new Table (see Supp. Table S4).
34
35

36 **3. SNP Discovery & Table S2: Pertinent to point #2, the authors report 53.5% novel**
37 **mutations in their SNP discovery effort when compared to HapMap. Is this in**
38 **comparison to Hapmap Phase II with 4 populations? This comparison should be done**
39 **with Phase III. And, the claim for novelty should be put in a frequency context. Are the**
40 **novel SNPs singletons in isolated populations? A table summarizing the genotype**
41 **frequencies(rather than allele frequencies) listed by population is needed to put the data**
42 **in a frequency and population context.**
43
44

45
46 In the same line of the previous response, there are less SNPs in Phase III than in Phase II.
47 The 53.5% of novel mutations corresponds to what is referenced today in dbSNP (every SNP
48 in HapMap is included in dbSNP). Since the time of submission of our paper, some of the
49 novel SNPs have been now referenced in the dbSNP database, so we have now 47.2% of
50 novel mutations in our paper. These have been submitted to the dbSNP database. Concerning
51 the novel SNPs that are found as singletons, there are neither particularly restricted nor more
52 present in particular/isolated populations.
53
54

55 As suggested by the reviewer, we have now added a table (Supp. Table S4) summarizing the
56 genotype frequencies of each SNP in our 3 continental populations and added the genotype
57 frequencies given by HapMap Phase II and III to put our results in a frequency and population
58 context.
59
60

1
2
3 **4. Nonsynonymous mutations & fitness (table 2): the authors use the Polyphen**
4 **algorithm to predict fitness effects of NS SNPs. Bioinformatic predictions of fitness are**
5 **highly inaccurate and this data analysis is unlikely to be very meaningful. Very subtle**
6 **amino acid changes can have dramatic effects on function. Large amino acid changes on**
7 **the protein surface in unconstrained areas can have no effect on function.**
8
9

10 Most bioinformatic predictions of fitness are based on 3D protein structure, for the few genes
11 whose protein structure has been characterized, and on the conservation of polymorphic sites
12 across numerous species and gene paralogs. These methods have been shown to present a
13 high false negative rate, i.e. they are highly conservative. Consequently, those mutations that
14 are predicted to be probably damaging have a high probability to be true positives, while
15 many mutations predicted as benign have good probability to be damaging. Importantly, as
16 said by the reviewer, these algorithms have low power to predict the exact impact of a
17 mutation on protein activity or stability, but rather indicate the relevance of the site of interest
18 for survival, because they are based conservation through evolution. We used the PolyPhen
19 algorithm to give some clues about the putative functional role of each non-synonymous
20 mutation. While in the first version of the manuscript we used the PolyPhen v1, we have now
21 used the recently released updated version of the algorithm (PolyPhen v2) (Adzhubei et al.,
22 2010, Nature Methods). PolyPhen-2 achieved true positive prediction rates of 92% on
23 HumDiv dataset. This updated version has been shown to provide a better confidence than
24 PolyPhen v1 (82%) and than any other predictive tool (Adzhubei et al., 2010, Nature
25 Methods). Indeed, in agreement with the reviewer's comment, some polymorphisms that were
26 predicted to be benign by Polyphen v1 now become possibly or probably damaging with
27 Polyphen v2. However, in our paper, the most important point supporting the somehow
28 deleterious effects of non-synonymous mutations does not come from the PolyPhen
29 predictions but instead from their observed frequencies *in natura*, which are all at very low
30 frequency except one in *IFNGR2* (interestingly predicted as being benign). More generally,
31 this result is also observed at the genome-wide level, since mutations predicted to be
32 damaging by Polyphen are generally found at lower population frequencies than mutations
33 predicted as being benign (Ng et al., 2009 Nature).
34
35
36
37
38
39

40 We have now updated the PolyPhen predictions using PolyPhen v2 and clarified the actual
41 meaning of these predictions, which are expected to be more accurate (Material and Methods
42 page 8, and Results pages 11-12). Because PolyPhen is widely used by the scientific
43 community, we still believe that our results may provide some comparative data to scientists
44 interested in predictive methods.
45
46

47 **5. Purifying selection of IFNg: An important finding of the paper is that IFNg has**
48 **evidence of purifying selection compared to IFNGR1 and IFNGR2 (Figs 2, 3). This data**
49 **looks solid and interesting and includes reference to a CA microsatellite and SNP 874**
50 **which have been partially functionally analyzed previously. Although the functional**
51 **effect of these polymorphisms on IFNg production remains incompletely characterized,**
52 **the connection of the data in this manuscript to this locus is suggestive of a biologically**
53 **meaningful genetic observation. One major question is whether the magnitude of effect**
54 **of the purifying selection (omega value and DIND test) are relatively high compared to**
55 **other genes. Although there is an interesting difference compared to IFNGR1 and**
56 **IFNGR2, how does this compare to other immune genes? Or other gene classes?**
57 **Although this could partially be treated in the Discussion, it could also be addressed**
58 **with analysis of HapMap data to emphasize the relative importance of the IFNg finding.**
59
60

1
2
3 We thank the reviewer for this useful comment. To put our results on a broader genome-wide
4 context of purifying selection, HapMap data is not suitable because it is based on genotyping
5 data and we need full resequencing data. In this case, 1,000 genomes data could be useful but
6 the low coverage of most of the pilot data available does not allow proper genotype calls (see
7 detailed explanation in point 2). To circumvent this limitation, we have now used the genome-
8 wide sequence dataset of 11,624 genes provided by Celera (Bustamante et al., 2005, Nature)
9 (see again point 2 for detailed explanation). This dataset, which has been already specifically
10 used for the genome-wide detection of purifying selection, provides reliable data on both
11 divergence and polymorphism of silent and non-synonymous sites. Among the 7,557 genes
12 presenting at least one nonsynonymous variant, we found that only 7.7% exhibit an ω value
13 (indicating purifying selection) lower than that observed for our *IFNG*. When restricting the
14 analyses to genes classified as being involved in immune system process, only 10.3% of them
15 presented ω values lower than that observed for *IFNG*.
16
17
18
19

20 We have now added in the revised version of our paper this detailed analysis (see Material
21 and Methods pages 7, Results pages 12-13 and Discussion page 19), where we integrate our
22 results in the more general context provided by these 11,624 genes.
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Referee: 2

Manuscript Review for the Journal Human Mutation Title: Evolutionary genetics evidence of an essential, non-redundant role of the IFN - γ pathway in protective immunity Manuscript number: humu-2010-0462 Authors: Manry J et al. In this manuscript the authors use a number of classical and well-established, as well as more recent powerful tests applied to DNA sequence diversity results at genomic loci encoding three specific genes in the IFN- γ pathway to uncover and distinguish the effects of natural selection. The choice of the genomic loci is well-based, because of the role of the gene products in innate immunity and adaptive cell-mediated immunity against intracellular pathogens. These have been major pathogens affecting human populations, and susceptibility or protection almost certainly has had an overarching effect on reproductive success throughout human evolutionary history. Three genes in the pathway (IFNG, IFNGR1, and IFNGR2) were re-sequenced in 186 individuals from a number of populations. After confirming Hardy-Weinberg equilibrium for detected SNPs, and after haplotype reconstruction a number of tests of neutrality and evolutionary selection were carried out, and the results analyzed and compared. Some of these tests are classical, but cannot readily distinguish effects of selection from those of population demographic events, whereas others which are based on re-sequencing and other approaches are more useful in making this distinction. More particularly, in order to correct for the mimicking effects of demography, the authors incorporated demographic models based on re-sequencing data of non-coding regions in different samples from comparable populations reported in the literature. Tests included neutrality statistics, such as Tajima's D, or Fu and Li's D and F, as well as Fay and Wu's H. In addition, tests considering both inter-species, and within-species diversity were conducted using the McDonald-Kreitman Poisson Random Field (MKPRF) test. Perhaps the most powerful and enlightening analysis in this study, looks for evidence of recent positive selection using DIND (Derived Intra-allelic Nucleotide Diversity), based on the levels of nucleotide diversity associated with haplotypes in the ancestral, versus the derived alleles. This test is based on the presumption that the derived allele under positive selection which reaches high population frequencies should present lower levels of nucleotide diversity at linked sites than expected (excluding singletons and doubletons). In addition, long-range haplotypes were sought. For amino-acid-altering mutations, predicted fitness effects were also examined, based on predicted protein structure or sequence conservation information. The bottom line findings from these results and analyses, were that the IFNG gene shows evidence of strong purifying selection against non-synonymous variants, consistent with intolerance to disruption in the function of this gene product, whereas the other two genes in the pathway examined, seemed to have evolved under more relaxed selective constraints. The evidence for possible population-specific positive selection is presented, but somewhat weaker. Overall, this is a thorough, thoughtful, elegant, and significant scientific contribution that is interesting and important from several points of view:

- Biologically interesting, in terms of the genes examined and the innate immune pathway.

- A tour de force and important template for using re-sequencing and combining multiple tests and approaches to unravel the relative influence of selection and demographic history in shaping human genomic diversity at loci of interest.

We thank the reviewer for these general comments and suggestions.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

I think that the manuscript could be somewhat improved with attention to the following points:

- While it is laudable that the authors have based their analyses and inferences on their own re-sequencing data, it is not clear whether they are taking advantage of additional data now available in the public domain from the 1000 Genomes Project, which could possibly strengthen and sharpen their conclusions.

We entirely agree with the reviewer in that the use of genomewide resequencing datasets (e.g. 1000 genomes, Celera, etc) is very useful for comparative purposes, especially when it comes to detect natural selection. The 1000 genomes project is in a pilot phase that includes 3 distinct approaches (Durbin et al, 2010, Nature). The first is a trio project where 2 trios have been whole-genome sequenced at high coverage, a sample size that is too low to provide with reliable comparisons in the context of our study. The second is an exon-targeted resequencing project where 900 genes have been sequenced at high coverage in 697 individuals. In this case, the number of exploitable genes is much lower than that of other public datasets (see Celera dataset discussed below), especially when it comes to make comparisons among different functional classes as the reviewer suggests. The third project is a low-coverage whole-genome sequencing in 179 individuals. Although the use of the third project of the 1000 genomes project is seducing, there are a number of reasons for which its use is not appropriate in the context of our study: (i) the sequence coverage varies among populations (see Suppl Fig 2 in Durbin et al., 2010 Nature), which complicates comparisons of neutrality statistics among populations, (ii) the variation in coverage along the genome strongly limits comparisons between genes within the same population, (iii) the low coverage of most of this dataset (in average 3.6x) limits the detection of low frequency variants, which in turn, represent the substrate to detect and estimate the intensity of purifying selection (i.e. one aim of our study). Indeed, it has been shown that, because non-synonymous variants are generally found at low frequencies (<5%), this dataset has reduced power to discover variants in this range, and therefore alter interpretation as to selection pressures (Durbin et al, 2010, Nature). In this view, we decided to use the Celera dataset, which contains 11,624 genes that have been fully resequenced (Bustamante et al., 2005, Nature), for the following reasons: (i) it is, by the time being, the resequencing project providing the largest number of resequenced genes (11,624); (ii) genes have been resequenced using standard PCR-based techniques, thereby excluding any of the limitations introduced by next-generation sequencing (coverage variation), and (iii) this dataset has been already specifically used for the genomewide detection of the intensity of purifying selection (Bustamante et al., 2005, Nature), constituting therefore a perfect comparative dataset in the context of our study.

In our revised version of the manuscript, we now use the Celera genome-wide dataset for comparative analyses (see Material and Methods page 7, Results pages 12-13 and Discussion page 19). For frequency comparisons, in turn, we have now used the HapMap Phase III data, obviously for those SNPs that have been genotyped by HapMap III (see Supp. Table S3). We have also complemented this comparison with a new Table (see Supp. Table S4).

- The authors could consider using ASD (allele sharing distance) for population demographic history, as it has a number of advantages related to potential biasing inherent in F_{ST} .

1
2
3 We agree with the reviewer than methods such as ASD can be helpful to detect population
4 relationships and admixture. However, the reason by which we did not use these methods is
5 that the aim of this paper is not to provide data on population demographic scenarios. In
6 addition, the analyses we performed using F_{ST} were not intended to enlighten population
7 demographic history but instead provide a background empirical distribution to better
8 understand the effects of natural selection. Indeed, the only reason by which we need a
9 demographic model is to correct the mimicking effects that demography and selection have on
10 the patterns of genetic diversity. To this end, we used the demographic scenario given by
11 (Voight et al. 2005, PNAS) determined using non coding regions, which are expected to be
12 more appropriated for demographic studies. Consequently, to detect robust signatures of
13 natural selection in our genes, we performed neutral simulations incorporating this
14 demographic model. In this respect, the Voight's model does not use F_{ST} to infer any
15 population demographic parameter. Furthermore, to avoid biases associated with the use of a
16 single demographic model (with its own methodological strengths and weaknesses), we have
17 now used in parallel a second demographic model that also used data from independent
18 noncoding regions (Laval et al., 2010, PLoS One). Interestingly, all tests for selection that
19 were significant under the Voight's model remain significant under the Laval's model. The
20 only discrepancy was the Fay and Wu's H statistics for *IFNG* in the African population,
21 which was found to be sensitive to intercontinental migration rates. This additional
22 information has now been added, and accordingly discussed, in the Material and Methods
23 section (page 7), Results (pages 14-15) and Discussion (page 20). We thank the reviewer for
24 this comment, which has allowed us to improve our understanding and interpretations of our
25 results.
26
27
28
29
30

31
32 **- The authors describe an interesting result with respect to SNP density. A higher SNP**
33 **density was found in African, than in the two Eurasian population groups which were**
34 **re-sequenced. Once again, an examination of the 1000 Genomes data would be helpful**
35 **here, although this would probably simply confirm the finding. However, it might allow**
36 **a more quantitative analysis and inference, to make this finding more significant and**
37 **far-reaching, and enable discussion and interpretation that is more confident. For**
38 **example, in the Results, the authors indicated that even if most of these population-**
39 **specific SNPs were found as singletons, a number of them display minor allele**
40 **frequencies exceeding 5%, as indicated in the Supplementary Table. This specific**
41 **finding could be solidified and consolidated with additional data from the 1000 Genomes**
42 **Project.**
43
44

45
46 We agree with the reviewer in that obtaining a higher SNP density in Africa compared to
47 Europe and Asia is expected and it is indeed confirmed by several studies (for an exhaustive
48 review, see Campbell & Tishkoff, 2010, Curr Biol). Also, we agree with the reviewer in that
49 comparing our data with the final phase of the 1000 Genomes project could allow a more
50 quantitative view of our results, in a genome-wide context. However, as described in our
51 response to the first point raised by the reviewer, the 1000 Genomes project in its pilot phase
52 does not allow such a quantitative analysis, because the sequencing coverage of African and
53 non-African populations is different. Specifically, the power of SNP discovery is increased in
54 European-Americans (CEU) with respect to Yoruba (YRI) and East-Asians (CHB+JPT).
55
56

57
58 We are now clearly mentioning this in the main text (Results page 11).
59

60 **- The examination of length variation at the +875 (CA)_n microsatellite within IFNG, is**
perhaps the least convincing and unimportant part of this manuscript. One can consider

1
2
3 **omitting it without much loss of the otherwise elegant and strong manuscript. In fact, I**
4 **think that this is a weakness, especially given the problematic nature of understanding**
5 **the evolution of microsatellite diversity.**
6
7

8 We entirely agree with the reviewer that the section related to the microsatellite is certainly
9 the less strong and the manuscript could perfectly “survive” without it. However, there is a
10 wide literature concerning the *IFNG* microsatellite, and many scientists are interested in this
11 topic beyond the aspect of natural selection. Therefore, we prefer to keep this extra
12 observation, more as descriptive data for those interested in this, than as an evidence of
13 natural selection. We think that this result can be interesting for immunologists interested in
14 the mechanisms of regulation of the production of IFN- γ .
15
16

17 **- A short description of the structure of the gene in the Introduction would help orient**
18 **the reader (here I am referring to chromosomal location, size, exons). This is**
19 **particularly true for the non-initiated reader.**
20
21

22 We added a Supp. Figure S1 to illustrate chromosomal location, size and exons. We also
23 describe the 3 genes and the proteins they encode in the revised version of the Introduction
24 (see page 4).
25
26

27 **- The authors should indicate the total size that was sequenced.**
28

29 The size of each sequenced region is now given in Supp. Table S2.
30
31

32 **- The GenBank number for each reference to sequence that was used should be**
33 **provided.**
34

35 The GenBank number is now added in Supp. Table S2.
36
37

38 **- A simple linkage disequilibrium (LD) Figure for each gene within each population**
39 **would be helpful to demonstrate high degrees of LD among SNPs. This is of course very**
40 **simple to create using Haploview.**
41
42

43 We have now added a new Supplementary Figure S3 to illustrate the LD between each SNP
44 with Minor Allele Frequency >1% for each gene within each population (see also Material
45 and Methods, page 6) and Results (page 15).
46
47
48

49 **- The ancestral state of SNPs would be of interest in some cases for which the haplotype**
50 **tree is not rooted.**
51
52

53 We added the ancestral and derived alleles of each SNP in the Supp. Figure S5.
54
55

56 **- One would assume that novel SNPs found in this study have been submitted to the**
57 **dbSNP database.**
58

59 We have now submitted the novel SNPs to the dbSNP database, and are waiting for the
60 corresponding #rs numbers.

EDITORIAL BOARD'S COMMENTS

This paper has several interesting observations stemming from a solid statistical analysis of resequencing data applied to sequence diversity results at specific genomic loci to uncover and distinguish the effects of natural selection and demographic history in shaping human genomic diversity at loci of interest. As such the paper is of interest yet it would benefit from proper consideration and handling of the reviewers' comments. In agreement with the latter we strongly suggest that the authors do a more thorough analysis of publicly available datasets and avoid possible biases from an unusual selection of samples for their dataset. In addition, the authors do not fully articulate the significance of their findings. In the HapMap age (not to mention the 1000G project), population geneticists can finally make more genome-wide statements regarding selective pressure on a candidate region. Ultimately we would like to know whether IFN γ shows evidence of a different level of purifying selection than other immune genes" the authors only partially get to that answer and they fill that out much more thoroughly with publicly available databases and a modest amount of work. Such additional substance would raise the impact of this contribution.

In revision, since the format is being changed, please carefully indicate in your response exactly what was added/removed in the revision process.

First of all, we thank the two reviewers as well as the editorial board for both the comments and suggestions as well as for the "upgrade" to research Article. In the revised version of the manuscript, we have followed all reviewers' and editorial's suggestions, which in our opinion have greatly contributed to improve the clarity and the quality of the manuscript. More generally, the revised version of the manuscript has been considerably changed and new figures and tables are now provided. The corresponding changes are now highlighted in red in the revised version of the manuscript.

In brief,

1. **Population choice of samples:** we now explain carefully the criteria and rationale of our population choice and study design, and showed that the presence of minorities in our sample collection does not influence our claims on selection (see detailed response to point 1 of Reviewer 1, and pages 5 and 18 of the revised manuscript, and the new Supp. Table S1).
2. **Use of public genome-wide datasets:** We entirely agree with both the reviewers and the editorial board in that the use of genomewide datasets (e.g. HapMap, 1000 genomes, Celera, etc) is very useful for comparative purposes. In the revised version of the manuscript, we are now comparing our data with genome-wide datasets, obviously when the data are comparable. Specifically, for frequency comparisons purposes (see reviewer 2), we are using both HapMap Phase II and III. Note however that SNP density is greater in HapMap Phase II (~3.1 million genotyped SNPs) than in HapMap Phase III (~1.4 million genotyped SNPs). We are therefore using now HapMap Phase III data for those SNPs that have been genotyped in this dataset. For tests regarding the detection of natural selection, one has to use datasets that are based on full resequencing (e.g., 1000 genomes, Celera dataset, etc), to circumvent the limitations associated with the nature of the HapMap data (i.e. SNP genotyping which is subject to ascertainment biases). Although the use of the 1000 genomes project is seducing, there are a number of reasons for which it is hardly comparable with our data: (i) the sequence coverage varies among populations (see Suppl Fig 2 in Durbin et al., 2010, Nature), which complicates comparisons of neutrality statistics among

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

populations, (ii) the variation in coverage along the genome strongly limits comparisons between genes within the same population, (iii) the low coverage of most of this dataset (in average 3.6x) limits the detection of low frequency variants, which in turn, represent the substrate to detect and estimate the intensity of purifying selection (our study). Indeed, it has been shown that, because non-synonymous variants are generally found at low frequencies (<5%), this dataset has reduced power to discover variants in this range, and therefore alter interpretation as to selection pressures (Durbin et al, 2010, Nature). In this view, we decided to use the Celera dataset, which contains 11,624 genes that have been fully resequenced (Bustamante et al., 2005, Nature), for the following reasons: (i) it is, by the time being, the resequencing project providing the largest number of resequenced genes (11,624); (ii) genes have been resequenced using standard PCR-based techniques, thereby excluding any of the limitations introduced by next-generation sequencing (coverage variation), and (iii) this dataset has been already specifically used for the genomewide detection of the intensity of purifying selection (Bustamante et al., 2005, Nature), constituting therefore a perfect comparative dataset in the context of our study.

We now discuss the results obtained for the *IFNG* in a much wider genome-wide context. For all these changes, see Material and Methods pages 7, Results pages 12-13 and Discussion page 19, as well as Supp. Table S3 and Supp. Table S4.

3. **Microsatellite:** The reviewer 2 suggests omitting the section related to the IFNG microsatellite, and we agree that the manuscript could perfectly “survive” without it. However, there is a wide literature concerning the IFNG microsatellite, and many scientists are interested in this topic beyond the aspect of natural selection. Therefore, we prefer to keep this extra observation, more as descriptive data for those interested in this, than as an evidence of natural selection. We think that this result can be interesting for immunologists interested in the mechanisms of regulation of the production of IFN- γ . Moreover, note that reviewer 1 comments the interest of using the microsatellite to provide some functional/biological meaning of our results on natural selection.
4. **Other points:** in addition to your requests, and those of the reviewers, new analyses have been done including:
 - We have used the most updated, and recently released, version of the Polyphen algorithm, which has been shown to have a highly accurate predictive power (PolyPhen v2, Adzhubei et al., 2010, Nature Methods) (see Material and Methods page 8, and Results pages 11-12)
 - We now correct our selection tests for two demographic models (instead of only one), see Material and Methods section (page 7), Results (pages 14-15) and Discussion (pages 20)
 - We have also taken advantage of genome-wide data from HapMap to perform another test of selection (XP-EHH, which is based on the degree of long range haplotype homozygosity). We particularly thank the reviewers and the editorial board to “push” us to go further in the genome-wide analyses since we have obtained an interesting result regarding the IFNGR1 gene in Africa (pages 8, 16, 22-23).
 - Because one of the reviewers asks for a table reporting the genotype information per individual, we have had to create an excel file (it is impossible to do as a .doc file) which is being sent as a separate email attachment to the editorial office.

We hope that we have now satisfied both the reviewers’ and the editorial board’s comments and that the manuscript is suitable for publication in *Human Mutation*.

MANAGING EDITOR COMMENTS:

Please respond to these points under "Response to Managing Editor", otherwise the final decision could be delayed.

1) Please include the OMIM accession number for the genes discussed, at first mention in the Introduction.

--Human gene symbols must be in all caps-italics.

--Please ensure that you use HUGO HGNC-approved gene symbols. Common gene symbol aliases may also be used at first mention. OMIM entries do not always feature the approved symbol prominently. Verify gene symbols at

<http://www.genenames.org/>

We added the OMIM number of the 3 genes at first mention in the introduction (see page 4).

2) The mutation nomenclature must follow the format indicated in the Author Instructions (see the website <http://www.hgvs.org/mutnomen/> and also the nomenclature checklist at <http://www.hgvs.org/mutnomen/checklist.html>). Things to watch for:

a-If there is a dbSNP accession number, then descriptions like rs123456:A>G are acceptable.

The mutation nomenclature follows the format indicated in the Author Instructions, and also the nomenclature checklist.

b-Mention the GenBank reference sequence and version number for the genes studied (i.e., include the decimal point following the accession number in the sequence record) in the Materials and Methods and as a footnote to the relevant tables.

We mentioned the GenBank reference sequence and version number for the 3 genes studied in the Material and Methods (Page 6) and in Supp. Table S2.

c-Clearly indicate in the Methods text and tables that the DNA mutation numbering system you follows the journal's approved nomenclature. For example:

"Nucleotide numbering reflects cDNA numbering with +1 corresponding to the A of the ATG translation initiation codon in the reference sequence, according to journal guidelines (www.hgvs.org/mutnomen). The initiation codon is codon 1."

We better described the journal's approved nomenclature (page 5).

d-Authors are advised to check sequence variant descriptions using the Mutalyzer program

(<http://www.LOVD.nl/mutalyzer/>) - using batch mode, all variants for a gene can be analyzed at once.

We checked sequence variants descriptions using the Mutalyser program.

e-Please verify that the mutations reported (especially novel ones) have been or will be submitted to an existing locus-specific database for the genes involved. Visit the HGVS-LSDB list to search for databases: <http://www.hgvs.org/dblist/glsdb.html>

1
2
3
4 Novel SNPs have been now submitted to the dbSNP database, and we are waiting for the
5 corresponding #rs numbers.
6
7

8 **3) On resubmission:**

9 **a-Please double check the author names and affiliations carefully. These are often a**
10 **source of typographical errors.**
11

12 Done
13

14
15 **b-An unformatted Title Page (with corresponding author contact information and other**
16 **affiliations), Abstract (180-200 words max), Key Words, Main Text, References, and**
17 **Figure Legends should be combined into one file for the manuscript and submitted as a**
18 ***.doc file.**
19

20 **--The text should be made 12 point double-spaced throughout.**
21

22 Done
23

24
25 **c-Figures for main article must be submitted as separate files with high resolution (at**
26 **least 200 dpi) as *.tif or *.eps format only.**

27 **--For color figures in print: submit two files for each color figure: one in CMYK color**
28 **space and one in RGB color space (with the true color you wish to have published). If**
29 **you cannot provide CMYK color e-files for the print version, please let me know. (see**
30 **below regarding color costs)**
31

32 **--For online-only color figures, please only submit one RGB color space file.**
33

34 Only Figure 3 is submitted with colours, in two separate files (CMYK and RGB).
35

36 **d-Tables must be submitted individually as separate *.doc files (with their titles and**
37 **legends included). Please use the MS Word table format if possible. Excel (*.xls) files**
38 **should not be submitted.**
39

40 **--Do not use custom paper sizes - only use Letter or A4.**
41

42 Table 1 and Table 2 are provided as separate *.doc files, with their titles and legends.
43

44
45 **e-Any Supporting Tables or Figures should be named and cited from the text as follows:**
46 **'Supp. Table S1' and 'Supp. Figure S1' (see below).**

47 **--If possible, do not use custom paper sizes - Letter or A4 are preferred.**
48

49 Done
50

51
52 **f- Supporting Figures and Tables should be prepared in a single MS Word *.doc file**
53 **labeled 'Supp_Mat', with Figures preceding Tables. Each table/figure should be**
54 **accompanied with its legend.**
55

56 All Supporting Tables and Figures have been merged into a single *.doc file. However, the
57 Supp. Table S4, which contains an excessive number of columns to be included in a standard
58 format, is provided as a *.xls file. In addition, this file in .xls format will be highly useful for
59 the readers (possibility of filtering, data mining, etc)
60

1
2
3 **3) Please check to see that the references follow the journal's standard format and are**
4 **cited properly. See our online Author Instructions**

5 [http://onlinelibrary.wiley.com/journal/10.1002/%28ISSN%291098-](http://onlinelibrary.wiley.com/journal/10.1002/%28ISSN%291098-1004/homepage/ForAuthors.html)
6 [1004/homepage/ForAuthors.html](http://onlinelibrary.wiley.com/journal/10.1002/%28ISSN%291098-1004/homepage/ForAuthors.html)
7
8

9 We checked that the references follow the journal's standard format and are cited properly.
10

11 **4) Figures 1 and 2 currently require color in the print version in order to be fully**
12 **understood because the color "red" is mentioned and must be identified. In Figure 1, it**
13 **could be re-drawn to use black or grey, thus eliminating need for color. Figure 2 may**
14 **still require color in order to be understandable unless different levels of gray/white are**
15 **used.**
16
17

18 Figures 1 and 2 are now in black and white.
19

20
21 **As noted in our Author Instructions, there are no page charges for publication in**
22 **Human Mutation but there are costs associated with publication of color images in**
23 **print: \$500 USD per printed color page. Alternatively, there is an option of publishing**
24 **the color images in black-and-white in the print article and in color online, at no cost to**
25 **you - but no information must be lost in a conversion to b/w from color for the print**
26 **version. Please confirm your preference in reply.**
27
28

29 We prefer Figure 3 to appear in colors in both the online and the printed versions.
30

31
32 **5) Human Mutation can accommodate researchers funded by agencies requiring open**
33 **access publication. More information on Wiley-Blackwell's policy is available at:**

34 <http://olabout.wiley.com/WileyCDA/Section/id-406241.html>
35

36 **Human Mutation abides by the NIH Mandate. If your work was funded by the NIH, be**
37 **sure to include a grant number in an "Acknowledgments" section of the manuscript**
38 **right before the References. Visit this site for more information:**

39 <http://www.wiley.com/go/nihmandate>
40
41

42 **6) IMPORTANT INFORMATION REGARDING PREPRINTS**

43
44 **a-Human Mutation is now publishing online preprints of accepted manuscripts prior to**
45 **typesetting and page proof corrections. It is therefore crucial that you revise your**
46 **manuscript carefully so that errors (typographical and grammatical) are corrected**
47 **BEFORE the final accepted manuscript is posted online. The accepted preprint version**
48 **will remain online until the corrected proofs are received and the typeset manuscript is**
49 **finalized. At that time, the preprint version will be replaced with the final, typeset**
50 **version online, in Early View.**
51
52

53
54 **b-It is essential that you submit a copyright transfer agreement (CTA) upon submission**
55 **of your revised manuscript. This will avoid delay in publication of your article upon**
56 **acceptance. If possible, the CTA must be signed by the corresponding author and should**
57 **be signed by all contributing authors if practical. All authors must be made aware of the**
58 **CTA and the rights it conveys to them.**
59

60 The CTA can be found here:

www.wiley.com/go/ctaus

1
2
3 **and must be filled out completely, including the article title and manuscript number.**
4 **Please fax it to this number: (201) 748-6091.**
5
6

7 **Submission of a CTA does not guarantee acceptance in the journal, but it will facilitate**
8 **rapid online publication of your paper if it is accepted.**
9

10 The CTA has now been faxed.
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

1
2
3 1 **Evolutionary genetics evidence of an essential, non-redundant role of the IFN- γ pathway in**
4
5
6 2 **protective immunity**
7
8 3

9
10 4 Jérémy Manry ^{1,2}, Guillaume Laval ^{1,2}, Etienne Patin ^{1,2}, Simona Fornarino ^{1,2}, Magali Tichit ³,
11
12 5 Christiane Bouchier ³, Luis B. Barreiro ⁴, Lluís Quintana-Murci ^{1,2}
13
14 6
15 7
16 8

17
18
19
20 8 ¹Institut Pasteur, Human Evolutionary Genetics, Department of Genomes and Genetics, F-75015
21
22 9 Paris, France; ²Centre National de la Recherche Scientifique, URA3012, F-75015 Paris, France ;
23

24 10 ³Institut Pasteur, Plate-forme Génomique, Pasteur Genopole, Paris, France; ⁴Department of
25
26 11 Human Genetics, University of Chicago, Chicago, USA
27
28 12

29
30
31 13 *Correspondence to Dr. Lluís Quintana-Murci, CNRS URA3012, UP Génétique Evolutive
32
33 14 Humaine, Institut Pasteur, 25 rue du Dr. Roux, 75724 Paris Cedex 15, France ; Phone :
34
35 15 +33.1.40.61.34.43 ; Fax :+33.1.45.68.86.39 ; E-mail : quintana@pasteur.fr
36
37
38 16
39
40
41 17

42
43 18 Short Title: Natural Selection acting on IFN- γ pathway
44
45
46 19
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 **ABSTRACT**
4

5
6 Identifying how natural selection has affected immunity-related genes can provide insights into
7
8 the mechanisms that have been crucial for our survival against infection. Rare disorders of either
9
10 chain of the IFN- γ receptor, but not of IFN- γ itself, have been shown to confer predisposition to
11
12 mycobacterial disease in patients otherwise normally resistant to most viruses. Here, we defined
13
14 the levels of naturally-occurring variation in the three specific genes controlling the IFN- γ
15
16 pathway (*IFNG*, *IFNGR1*, *IFNGR2*) and assessed whether and how natural selection has acted on
17
18 them. To this end, we resequenced the three genes in 186 individuals from sub-Saharan Africa,
19
20 Europe and East-Asia. Our results show that *IFNG* is subject to strong purifying selection against
21
22 nonsynonymous variants. Conversely, *IFNGR1* and *IFNGR2* evolve under more relaxed selective
23
24 constraints, although they are not completely free to accumulate amino-acid variation having a
25
26 major impact on protein function. In addition, we have identified signatures of population-
27
28 specific positive selection, including at one intronic variant known to be associated with higher
29
30 production of IFN- γ . The integration of our population genetic data into a clinical framework
31
32 demonstrates that the IFN- γ pathway is essential and non-redundant in host defense, probably
33
34 because of its role in protective immunity against mycobacteria.
35
36
37
38
39
40
41
42

43 **KEY WORDS:** *IFNG*, *IFNGR1*, *IFNGR2*, polymorphisms, natural selection, population genetics
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1 INTRODUCTION

2 Interferons (IFNs) are helicoidal cytokines that play a key role in innate and adaptive immune
3 responses. Most IFNs present an antiviral activity and are intercellular mediators able to
4 modulate several major biological functions, such as cell proliferation and differentiation, or
5 lymphocyte activation. IFNs are today classified into three types, on the basis of gene sequence
6 similarity, chromosomal location, and receptor specificity (see [Pestka et al., 2004] for an
7 extensive review). The first IFNs to be identified were classified as type-I IFNs (17 molecules of
8 IFN- α/β and related molecules) and signal through a ubiquitously expressed receptor composed
9 of two chains: IFN- α R1 and IFN- α R2. The last IFNs to be described are known as type-III IFNs,
10 and the three molecules that have been identified so far (IL28A, IL28B and IL29) activate the
11 same main signaling pathway as type-I IFNs but have evolved a completely different receptor
12 structure. Type-III IFNs act through a receptor composed of two chains, a type-III IFN-specific
13 IL28RA selectively expressed in certain cell types and the ubiquitously expressed IL10RB.
14 Finally, only one type-II IFN has been identified, IFN- γ , which presents a distinct sequence, role
15 and functions with respect to type-I and type-III IFNs [Schroder et al., 2004].

16 **The type-II IFN- γ binds to its own receptor made of the 2 transmembrane proteins, IFN-
17 γ R1 and IFN- γ R2, to both induce antimicrobial and antitumor mechanisms and to up-regulate
18 antigen processing and presentation pathways. More precisely, IFN- γ , which is produced mostly
19 by natural killer (NK) and T lymphocytes, orchestrates leukocyte attraction and directs growth,
20 maturation, and differentiation of many cell types, in addition to enhancing NK cell activity and
21 regulating B-cell functions such as immunoglobulin (Ig) production and class switching
22 [Schroder et al., 2004]. Consequently, IFN- γ plays a central role in innate immunity and in
23 adaptive cell-mediated immunity against intracellular pathogens and is the major macrophage-**

1
2
3 1 activating cytokine. Today, IFN- γ is used to treat chronic granulomatous disease [Todd and Goa,
4
5 2 1992], osteopetrosis [Key et al., 1992] and IL12/IL12RB1 deficiency [Filipe-Santos et al., 2006].
6
7

8 The IFN- γ protein, which is composed of 166 amino-acids including the signal peptide, is
9
10 encoded by the *IFNG* gene (MIM# 147570), which is located on chromosome 12 and is
11
12 composed of 4 exons (Supp. Figure S1). The *IFNGR1* (MIM# 107470) and *IFNGR2* (MIM#
13
14 147569) genes, containing both 7 exons (Supp. Figure S1), encode the two receptor subunits of
15
16 489 and 337 amino-acids including the signal peptide and are located on chromosomes 6 and 21,
17
18 respectively. Whereas *IFNGR1* is physically isolated from other IFN receptors, *IFNGR2* is
19
20 located in a cluster of 4 genes (*IFNARI*, *IFNAR2*, *IL10RB* and *IFNGR2*) that are all known to
21
22 interact with IFN proteins.
23
24
25
26

27 The evolutionary genetics approach has proven to be useful to increase our understanding
28
29 of the evolutionary forces that affect the human genome, providing an indispensable complement
30
31 to clinical and epidemiological genetics approaches [Akey, 2009; Barreiro and Quintana-Murci,
32
33 2010; Di Rienzo, 2006; Nielsen, et al., 2007; Quintana-Murci, et al., 2007; Sabeti, et al., 2006].
34
35
36 The aims of this study were to (i) identify the whole spectrum of population genetic variation,
37
38 based on a full resequencing scheme, in the three core genes involved in the IFN- γ pathway
39
40 (*IFNG*, *IFNGR1* and *IFNGR2*), and (ii) investigate, using an evolutionary genetics approach,
41
42 whether and how natural selection has targeted the type-II IFN system. Identifying the intensity
43
44 and type of natural selection exerted upon these 3 genes should help to better understand the
45
46 mode in which the different components of the type-II IFN system have contributed to host
47
48 defense. Likewise, the evolutionary dissection of *IFNG*, *IFNGR1* and *IFNGR2* should shed light
49
50 on how genetic variation at these genes may be involved in the current susceptibility to, and
51
52 pathogenesis of, infectious diseases.
53
54
55
56
57
58
59
60

1 MATERIALS AND METHODS

2 Population samples

3 Sequence variation for the *IFNG*, *IFNGR1* and *IFNGR2* genes was determined in 186 individuals
4 from sub-Saharan Africa, Europe and East-Asia (62 individuals per geographic region) from the
5 HGDP-CEPH panel [Cann et al., 2002]. Sub-Saharan African populations were composed of 19
6 Bantu from Kenya, 21 Mandenka from Senegal and 22 Yoruba from Nigeria; European
7 populations were composed of 20 French, 14 Italians, 6 Orcadians and 22 Russians; and East-
8 Asian populations were composed of 15 Han Chinese and 33 individuals from Chinese
9 minorities, 10 Japanese and 4 Cambodians. For a complete description of this HGDP-CEPH sub-
10 panel, see Supp. Table S1. This study was approved by the Institut Pasteur Institutional Review
11 Board (n° RBM 2008.06).

12 DNA Resequencing and SNP Discovery

13 For each gene, the totality of the exonic region and at least an equivalent amount of non-exonic
14 regions were sequenced, including intronic, 5' and 3' regions (Supp. Table S2). Nucleotide
15 numbering reflects cDNA numbering with +1 corresponding to the A of the ATG translation
16 initiation codon in the reference sequence, according to journal guidelines
17 (www.hgvs.org/mutnomen). The initiation codon is codon 1. In addition to the annotation of
18 coding DNA described above [den Dunnen and Antonarakis, 2000], we used another SNP
19 annotation of genomic DNA (referred to as “ATG position”) starting at the first nucleotide of the
20 ATG-translation initiation codon (+1) and including in the counting both coding and noncoding
21 nucleotides. All sequences were obtained using the Big Dye terminator kit and the 3730 XL
22 automated sequencer from Applied Biosystems. Sequence files and chromatograms were
23 inspected using the GENALYS software [Takahashi et al., 2003]. All singletons or ambiguous
24 polymorphisms were systematically reamplified and resequenced. We were unable to resequence

1 the first exon of *IFNGR2* because of technical reasons, most likely resulting from the very high
2 GC content of the region (73%). We used the NG_015840.1, NG_007394.1 and NG_007570.1 as
3 reference sequences for *IFNG*, *IFNGR1* and *IFNGR2*, respectively.

4 ***IFNG* +875 CA microsatellite genotyping**

5 We performed a standard PCR protocol to amplify the fragment using the True Allele PCR
6 Premix by Applied Biosystems with 20ng of genomic DNA, and using primer sequences
7 previously reported [Ding et al., 2008]. The mixture was then subjected to the PCR reaction for
8 15 min at 95°C, followed by 35 cycles of denaturation for 30 sec at 95°C, annealing for 1 min at
9 56°C, and extension for 1 min at 72°C, followed by a final extension of 10 min at 72°C. The
10 fluorescent dye-labelled PCR products were electrophoresed on an Applied Biosystems 3130XL
11 Genetic Analyser. The results were analyzed by Genemapper Analysis software 3.2.

12 **Statistical Analyses**

13 We checked the Hardy-Weinberg equilibrium for each SNP using Arlequin Software v3
14 [Excoffier et al., 2005]. We used the Haploview software [Barrett et al., 2005] to illustrate the
15 levels of linkage disequilibrium (LD) between each SNP for each gene. Haplotype reconstruction
16 was performed by means of the Bayesian statistical method implemented in Phase (v.2.1.1)
17 [Stephens and Donnelly, 2003]. We applied the algorithm five times, using different randomly
18 generated seeds, and consistent results were obtained across runs. The entire dataset was used to
19 perform a number of sequence-based neutrality-statistics, including Tajima's D , Fu & Li's D^* ,
20 Fu & Li's F^* , Fay & Wu's H , using DnaSP v5.1 [Rozas et al., 2003]. P -values for the various
21 neutrality tests were estimated from 10^4 coalescent simulations, performed using SIMCOAL 2.0
22 [Laval and Excoffier, 2004], under a finite-site neutral model and considering the recombination
23 rate of the tested region reported in HapMap Phase II [Frazer et al., 2007; International-HapMap-
24 Consortium, 2005]. Each of the 10^4 coalescent simulations was conditional on the observed

1
2
3 1 sample size and the number of segregating sites observed in each gene. To correct for the
4
5 2 mimicking effects of demography on the patterns of diversity, we considered **two previously**
6
7 **validated demographic models based on resequencing data of noncoding regions in a set of**
8
9 **populations similar to ours (i.e., African, European and Asian) [Laval et al., 2010; Voight et al.,**
10
11 **2005]. The main difference between these two demographic models is that the model of [Laval et**
12
13 **al., 2010] considers inter-continental population migration.**

14
15
16
17 To detect the effects of natural selection considering both inter-species divergence and
18
19 within-species polymorphism, we used the McDonald-Kreitman Poisson Random Field
20
21 (MKPRF) test [Bustamante et al., 2005; Sawyer and Hartl, 1992]. **We compared these MKPRF**
22
23 **results with a genome-wide dataset where 20 European-Americans and 19 African-Americans**
24
25 **have been resequenced at 11,624 genes by exon-specific PCR amplification [Bustamante et al.,**
26
27 **2005]. We used information on the number of divergent silent sites (d_S), and polymorphic silent**
28
29 **sites (p_S), divergent nonsynonymous sites (d_N) and polymorphic nonsynonymous sites (p_N) for**
30
31 **each gene. Divergent sites refer to positions that are different between the human and chimpanzee**
32
33 **lineages, whereas polymorphic sites refer to the situation in which the two alleles are segregating**
34
35 **in humans. In addition, we used the Gene Ontology database (<http://www.geneontology.org/>) to**
36
37 **extract genes involved in the biological process referred as to “immune system process”**
38
39 **(GO:0002376).**

40
41
42
43 To detect recent events of positive selection, we used the Derived Intra-allelic Nucleotide
44
45 Diversity (DIND) test based on the ratio $i\pi_A/i\pi_D$, where $i\pi_A$ and $i\pi_D$ are the levels of nucleotide
46
47 diversity associated with the haplotypes carrying the ancestral and the derived allele, respectively
48
49 [Barreiro et al., 2009]. The rationale of this test is that a derived allele under positive selection
50
51 that is at high population frequencies should present lower levels of nucleotide diversity at linked
52
53 sites than expected. Singletons and doubletons are excluded from this analysis. We also used tests
54
55
56
57
58
59
60

1 based on the levels of haplotype homozygosity, such as the Long Range Haplotype test [Sabeti et
2 al., 2002] and the Cross Population Extended Haplotype Homozygosity (XP-EHH) test [Sabeti et
3 al., 2007]. In addition, we assessed the levels of population differentiation for the entire SNP
4 panel, using the F_{ST} statistics derived from the analysis of variance (ANOVA) [Excoffier et al.,
5 1992]. To identify SNPs presenting extreme levels of population differentiation, we compared the
6 observed F_{ST} values at the level of individual SNPs in *IFNG*, *IFNGR1* and *IFNGR2* against the
7 F_{ST} distribution of 659,000 SNPs genotyped on the same subset of individuals of the HGDP-
8 CEPH, except for 5 individuals who were not genotyped [Li et al., 2008]. F_{ST} comparisons were
9 conditioned to SNPs presenting similar allele frequencies (i.e., similar expected heterozygosity).
10 Empirical P -values for each SNP at the 3 genes were estimated as previously described [Barreiro
11 et al., 2009].

12 The fitness status of all amino-acid-altering mutations (i.e., benign, possibly damaging
13 and probably damaging) was predicted using the Polyphen algorithm v2 HumDiv [Adzhubei et
14 al., 2010]. This method, which considers protein structure and/or sequence conservation
15 information for each gene, has been shown to be the best predictor of the fitness effects of amino-
16 acid substitutions [Williamson et al., 2005]. To independently assess the functional impact of
17 these mutations, we replicated the analyses using the Panther algorithm [Thomas et al., 2003]

18 To compute the *IFNG* gene tree, we used the GENETREE software [Griffiths and Tavaré,
19 1994], under a standard coalescent model. Since this model assumes no recombination, we
20 excluded 3 recombinant haplotypes to perform the analysis. We used the θ_w obtained for the
21 entire *IFNG* sequenced region and the mutation rate per gene per generation ($\mu = 1.98 \times 10^{-8}$) was
22 deduced from D_{xy} (0.0095), the average number of nucleotide substitutions per site between
23 human and chimpanzee calculated by DnaSP v5.1 [Rozas et al., 2003], with consideration that
24 the two species diverged 240,000 generations ago. Time, scaled in $2N_e$ units, with N_e the

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

- 1 effective population size, was converted into years by the use of a 25-year generation time and an
- 2 N_e value obtained as θ_w divided by 4μ .

For Peer Review

1 RESULTS

2 SNP Discovery in World Populations and Global Levels of Genetic Diversity

3 To assess the levels of full sequence-based diversity in the human *IFNG* and the 2 genes
4 encoding its receptor (*IFNGR1* and *IFNGR2*), we comprehensively resequenced the 3 genes in a
5 panel of 186 healthy individuals originating from 11 different populations from sub-Saharan
6 Africa, from Europe and from East-Asia. Each individual was sequenced for a total of 14.8 kb,
7 32.5% of which corresponded to exonic regions, the rest comprising intronic and promoter
8 regions (Supp. Table S2). Our population-based resequencing effort allowed us to identify 127
9 mutations, including 117 single nucleotide polymorphisms (SNPs), 2 insertions, 5 deletions and 3
10 duplications (Supp. Table S3). **Out of the 127 polymorphisms here reported (excluding the *IFNG***
11 **+875 (CA)_n microsatellite), 60 (47.2%) were novel and not previously reported in the dbSNP**
12 **database, and 105 (82.7%) have not been genotyped by the HapMap Consortium [Altshuler et al.,**
13 **2010; Frazer et al., 2007] (Supp. Table S3).**

14 The three genes, particularly *IFNG* and *IFNGR1*, displayed generally low levels of
15 nucleotide diversity per site (π): 4×10^{-4} , 4.7×10^{-4} and 7.2×10^{-4} for *IFNG*, *IFNGR1* and *IFNGR2*,
16 respectively. To compare these levels of nucleotide diversity with background genic expectations,
17 we used the SeattleSNPs database, which reports the sequence diversity of 327 genes involved in
18 inflammatory responses in similar human populations. *IFNG* and *IFNGR1* were found to fall in
19 the 15th percentile of genes presenting the lowest nucleotide diversity, whereas *IFNGR2* was
20 included in the 46th (Table 1). Population-wise, *IFNG* and *IFNGR2*, genetic diversity was higher
21 in Europeans relative to both Africans and East-Asians (Table 1). At the haplotype level, we
22 observed the expected picture of Africans displaying higher levels of haplotype diversity than
23 non-African samples for the 3 genes (Table 1).

1
2
3 1 With respect to SNP density, we observed more SNPs in Africa than in the two Eurasian
4
5 2 population groups, an observation that is compatible with public datasets based on genome-wide
6
7 3 genotyping data, such as the HapMap Phase II and III datasets [Altshuler et al., 2010; Frazer et
8
9 4 al., 2007], or whole genome sequencing data, such as the 1000 Genomes Project [Durbin et al.,
10
11 5 2010]. We found 75, 39 and 45 SNPs in Africa, Europe and Asia, including 53, 16 and 24
12
13 6 population-specific SNPs, respectively. Interestingly, even if most of these population-specific
14
15 7 SNPs are found as singletons (28, 12 and 15 in Africa, Europe and Asia, respectively), a number
16
17 8 of them display minor allele frequencies $\geq 5\%$ (Supp. Table S3). For those SNPs identified in our
18
19 9 study that have been genotyped by the HapMap Phase II and/or III, we compared their genotype
20
21 10 frequencies in our population panel with those of multiple populations worldwide [Altshuler et
22
23 11 al., 2010] (Supp. Table S4). In virtually all cases, population frequencies were highly comparable
24
25 12 at the level of the different continental populations.
26
27
28
29
30
31
32
33

34 **Putative Functional Consequences of Nonsynonymous Variants**

35
36 15 Among the 26 exonic mutations identified, 16 SNPs corresponded to nonsynonymous mutations:
37
38 16 10 in *IFNGR1* and 6 in *IFNGR2* (Table 2). All these nonsynonymous mutations were found at
39
40 17 frequencies lower than 5% but one, in *IFNGR2*, that displayed high population frequencies in the
41
42 18 three continental population groups, presenting a derived allele frequency ranging from 62% in
43
44 19 Asia to 88% in Europe. It is worth noting that no nonsynonymous mutations were observed at
45
46 20 *IFNG*. To give some clues on the potential functional effect of the non-synonymous mutations
47
48 21 we found, we predicted the functional effects of both nonsynonymous mutations fixed between
49
50 22 the human and the chimpanzee lineages and those that are polymorphic within humans, using the
51
52 23 **Polyphen v2 HumDiv algorithm** [Adzhubei et al., 2010]. This method, which considers protein
53
54 24 structure and/or sequence conservation for each gene, has been shown to be the best available
55
56
57
58
59
60

1 predictor of the fitness effects of nonsynonymous variants [Adzhubei et al., 2010; Williamson et
2 al., 2005]. Concerning nonsynonymous mutations that are fixed between the two species, we
3 found one in IFN- γ at position p.Val147Ala, two in IFN- γ R1 at positions p.Leu198Ile and
4 p.Ala366Val, and three in IFN- γ R2 at positions p.Thr76Met, p.Ser205Phe and p.Ala207Val. Five
5 of these fixed, divergent nonsynonymous substitutions were predicted to be benign, and one was
6 predicted to be **probably** damaging in IFN- γ R2 (p.Ser207Phe). Of the 16 nonsynonymous
7 mutations that are polymorphic in humans (Table 2), 9 were predicted to be benign. **The**
8 **remaining non-synonymous polymorphisms were predicted to be possibly or probably damaging.**
9 **It is worth noting that the sole nonsynonymous polymorphism with a frequency higher than 5% is**
10 **predicted to be benign.**

12 **Measuring the Intensity of Natural Selection in the Human Lineage**

13 To assess whether and how natural selection has operated on the human *IFNG*, *IFNGR1* and
14 *IFNGR2*, we first estimated the direction and strength of selection acting in the human lineage as
15 a whole. **To this end, we measured d_S and d_N , i.e. the number of silent and nonsynonymous fixed**
16 **differences between species (humans versus chimpanzee) as well as p_S and p_N , i.e. the number of**
17 **silent and nonsynonymous polymorphic sites observed within species (within humans). We used**
18 **the McDonald-Kreitman Poisson Random Field (MKPRF) test [Bustamante et al., 2005;**
19 **Bustamante et al., 2002; Sawyer and Hartl, 1992] in order to estimate ω (i.e., $\omega \propto \log[\theta_R/\theta_S]$),**
20 **which measures the selective pressure acting on amino-acid substitutions.** Under neutrality, ω is
21 not significantly different from 1. Lower values are consistent with selection against
22 nonsynonymous variants (strong purifying selection), whereas higher values reflect selection
23 favouring amino-acid changes (positive selection). *IFNG* presented a ω value significantly lower
24 than 1, indicating that this gene has evolved under the effects of strong purifying selection

1 (Figure 1A). We compared our data with a genome-wide resequencing dataset of 11,624 genes,
2 for which the d_S , d_N , p_S and p_N values are provided [Bustamante et al., 2005]. This study proposed
3 that the set of genes that are informative to detect the effects of selective constraints against
4 nonsynonymous variation are those displaying at least two variable non-synonymous sites (d_N+p_N
5 ≥ 2). However, because the *IFNG* gene does not fall into this category (i.e., $d_N+p_N=1$), we relaxed
6 this criterion by comparing the results of the ω parameter for *IFNG* with all genes displaying a
7 $d_N+p_N \geq 1$. Among the 7,557 genes falling into this category, we found that only 7.7% of them
8 exhibit a ω value lower than that observed for *IFNG*. When restricting the analyses to those genes
9 classified as being involved in “immune system process”, only 10.3% of them presented ω values
10 lower than that displayed by *IFNG*.

11 We next used the population selection parameter γ [Bustamante et al., 2005; Bustamante
12 et al., 2002; Gilad et al., 2003] to identify whether *IFNG*, *IFNGR1* and/or *IFNGR2* were subject
13 to selection operating on nonsynonymous mutations that are polymorphic in humans (i.e.,
14 segregating non lethal mutations). The parameter γ is negative if a gene displays an excess of
15 amino-acid polymorphisms within humans with respect to amino-acid divergence between
16 species (weak negative and/or balancing selection). Conversely, positive γ values reflect an
17 excess of amino-acid divergence with respect to amino-acid polymorphism (positive selection in
18 the human lineage) [Bustamante et al., 2005; Bustamante et al., 2002; Gilad et al., 2003].
19 *IFNGR1* presented a γ value significantly lower than 0 ($\gamma=-1.3$) (Figure 1B). This, together with
20 the fact that all nonsynonymous SNPs are found as singletons or doubletons, suggests that weak
21 negative selection maintains mutations causing amino-acid changes in *IFNGR1* at low population
22 frequencies because of their likely weakly deleterious effects on an individual’s fitness.

1 Intra-Species Sequence-Based Neutrality Tests

2 Intra-species sequence-based neutrality tests (i.e., Tajima's D , Fu and Li's D^* and F^* and Fay
3 and Wu's H) allowed us to evaluate whether the frequency spectrum of the 3 genes deviates from
4 expectations under neutrality and to detect therefore selection within human populations [Nielsen
5 et al., 2007; Sabeti et al., 2006; Wall, 1999]. Because these tests are known to be sensitive to the
6 mimicking effects that demography and selection have on the patterns of genetic diversity, we
7 considered a demographic model previously validated using a set of 50 unlinked noncoding
8 regions sequenced in a set of populations similar to ours (i.e., African, European and Asian)
9 [Voight et al., 2005]. This model considers a bottleneck in non-African populations starting
10 40,000 years ago in an ancestral population of 9,450 individuals, and an exponential expansion in
11 African populations. **In addition, we used an independent demographic model based on the
12 patterns of sequence diversity at 20 unlinked, non-coding regions in a set of populations from
13 Africa, Europe and Asia [Laval et al., 2010]. In contrast with the model of [Voight et al., 2005],
14 the latter does consider the occurrence of inter-continental population migration [Laval et al.,
15 2010].** Most tests indicated that the frequency spectrum of the three genes does not significantly
16 deviate from neutral expectations (Table 1). The exceptions were *IFNG*, for which we detected a
17 significant excess of high-frequency derived variants in Africa as attested by the significantly
18 negative values of Fay and Wu's H , and *IFNGR2*, for which we detected an excess of singletons
19 in Asia as attested by both Fu and Li's D^* and F^* tests (Table 1). **Using the model that considers
20 intercontinental migration [Laval et al., 2010], the significance of the Fay and Wu's H for *IFNG*
21 disappeared.**

23 Detecting Recent Events of Positive Selection

1
2
3 1 To identify recent events of positive selection acting on the three genes involved in the IFN- γ
4
5 2 pathway, we used the Derived Intra-allelic Nucleotide Diversity (DIND) test that makes
6
7 3 maximum use of resequencing data [Barreiro et al., 2009]. The rationale of this test is that a
8
9 4 derived allele under positive selection that is at high population frequencies should present lower
10
11 5 levels of nucleotide diversity at linked sites than expected. We applied the DIND test to our entire
12
13 6 dataset by plotting for all SNPs identified in the 3 genes, the ratio between the ancestral and the
14
15 7 derived internal nucleotide diversity ($i\pi_A/i\pi_D$) against the frequencies of the derived alleles
16
17 8 (Figure 2 and **Supp. Figure S2 when considering the models of [Voight et al., 2005] and [Laval et**
18
19 9 **al., 2010], respectively**). Our analyses identified a signature of positive selection targeting the
20
21 10 *IFNG* SNPs +874T (c.115-483A>T) and +5173G (c.*910A>G) in Europe (Figure 2B and Supp
22
23 11 Figure S2 B). Indeed, although the frequency of these two SNPs — which are found to be in
24
25 12 almost perfect linkage disequilibrium (LD) (**Supp. Figure S3 A-C**) with an r^2 value of 0.975 — is
26
27 13 very high in Europe (54.04%), only 2 internal haplotypes are observed: one defined by a
28
29 14 singleton (SNP +1201G [c.115-156A>G]), and the other accounting for the remaining 53.23% of
30
31 15 frequency. Our analysis also identified a significant signature of positive selection targeting
32
33 16 *IFNGR2* at SNP +23133A (c.413-209G>A) in Asia (Figure 2 I and **Supp. Figure S2 I**).

34
35
36
37
38
39
40
41 17 We next explored the extent of haplotype homozygosity surrounding the three genes, to
42
43 18 detect more recent events of positive selection. To this end, we used available genotype databases
44
45 19 (i.e., HapMap, HGDP-CEPH) [Frazer et al., 2007; Li et al., 2008] that contain a sufficient
46
47 20 number of SNPs over much larger physical distances than our resequencing data, a feature that is
48
49 21 needed to assess the levels of haplotype homozygosity. We did not detect any signature of recent
50
51 22 positive selection when using Long Range Haplotype (LRH)-based methods [Sabeti et al., 2002].
52
53 23 This could be due to the presence of recombination hotspots, which shorten the length of the
54
55 24 haplotypes, in the vicinity of the three genes. For example, *IFNG* is located less than 50kb away
56
57
58
59
60

1 from one of the strongest recombination hotspots of chromosome 12 [Myers et al., 2005].

2 However, when using the XP-EHH test, which detects alleles that have rapidly increased in
3 frequency in one but not all populations [Sabeti et al., 2007], we detected a significant XP-EHH
4 for *IFNGRI*, suggesting the action of positive selection in the African population (Supp. Figure
5 S4).

7 **Levels of Population Differentiation**

8 An alternative approach to detect population-specific events of positive selection is to calculate
9 genetic distances among populations, using the F_{ST} statistic [Excoffier et al., 1992; Weir, 1984].
10 Indeed, local positive selection is known to increase the levels of population differentiation with
11 respect to neutrally-evolving loci [Barreiro et al., 2008; Kreitman, 2000; Nielsen et al., 2007;
12 Sabeti et al., 2006; Voight et al., 2005]. We thus estimated the F_{ST} values for the 117 SNPs
13 identified in our study. To obtain a background expectation of genome-wide F_{ST} , we used the
14 genome-wide data of the HGDP-CEPH (~659,000 SNPs from [Li et al., 2008]) from the same set
15 of individuals we sequenced in this study. When plotting the F_{ST} values as a function of expected
16 heterozygosity for our 117 SNPs together with the background genome-wide expectations, 2
17 SNPs in *IFNGR2* (SNP +23133 [c.413-209G>A] and SNP +23501 [c.561+11G>C]) displayed
18 extreme levels of population differentiation, falling out of the 95th percentile of the global F_{ST}
19 distribution (Figure 3). Indeed, SNP+23133A reaches 37% in East-Asia while is absent in Africa,
20 and SNP+23501C reaches 32% in Europe while is virtually absent in East-Asia.

22 **Length variation at the *IFNG* +875(CA)_n microsatellite**

23 Because, based on the DIND results, we suspected that positive selection has targeted the *IFNG*
24 SNPs +874T (c.115-483A>T) and +5173G (c.*910A>G) in Europe, we took profit of existing

1
2
3 1 functional data on one of these SNPs to better understand the nature of the selective event.
4
5 2 Indeed, the +874T allele has already been documented to create an NF- κ B binding site, and to
6
7 3 lead to higher production of IFN- γ [Pravica et al., 1999; Pravica et al., 2000]. Moreover, it has
8
9 4 been shown that the +874T allele is associated with the 12 CA repeat-allele at the *IFNG*
10
11 5 +875(CA)_n (c.115-482CA11_18) microsatellite. We thus decided to assess the patterns of
12
13 6 microsatellite diversity associated with the +874A/T alleles. Among the 372 chromosomes we
14
15 7 genotyped at this microsatellite, we found 8 different alleles ranging from 11 to 18 CA repeats. In
16
17 8 contrast with previous observations proposing an absolute correlation between the 12 CA repeat-
18
19 9 allele at the microsatellite and the T allele at SNP +874 [Ding et al., 2008; Pravica et al., 2000],
20
21 10 we found individuals homozygous for the +874T allele, who were heterozygous for the
22
23 11 microsatellite. Indeed, at the haplotype level, we found 12 chromosomes, among the 112 carrying
24
25 12 the +874T allele, that were non-12 CA repeat (i.e. 13 or 15 CA repeats), and 13 chromosomes,
26
27 13 among the 260 carrying the +874A allele, that were 12 CA repeat (Supp. Table S5). It would be
28
29 14 now interesting to distinguish which of these two mutational events, i.e. the +874T allele or the
30
31 15 12 CA repeat, account for by the previously observed higher production of IFN- γ . More
32
33 16 importantly, we observed that the microsatellite diversity associated with the +874T allele was
34
35 17 much lower than that associated with the +874A allele (expected global heterozygosity 0.19 vs.
36
37 18 0.64, respectively). This observation further supported the notion that positive selection has
38
39 19 targeted the *IFNG* SNP +874T.
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

DISCUSSION

In this study, we sought to describe the levels of naturally-occurring variation in the three specific genes controlling the IFN- γ pathway (*IFNG*, *IFNGR1*, and *IFNGR2*) and to assess whether and how natural selection has acted on them. The first objective, the discovery of new polymorphisms, is optimal when sampled individuals come from many different ethnic origins, while the second objective, the identification of signatures of natural selection, requires many individuals per continental region. Our sample was designed to achieve both objectives, by including 186 individuals originating from multiple ethnic groups living in sub-Saharan Africa, Europe and Asia. However, a possible limitation of this population scheme is the potential presence of genetic structure among populations from the same continent, which are merged in our analyses, a situation that can lead to significant deviations of the allele frequency spectrum [Przeworski et al., 2005] and therefore create spurious signals of selection. To test this possibility, we performed an AMOVA [Excoffier et al., 1992] to estimate the fraction of the genetic variance of our dataset that is explained by genetic differences within a given population, among populations of a given continent, and among continental groups, and obtained values of 89.27%, 0.45% and 10.28%, respectively. The negligible, non-significant differentiation observed among populations from the same continent in our dataset (0.45%) is consistent with a genome-wide study conducted in the same individuals and populations, where population structure within continental regions was found to be limited [Li et al., 2008]. In addition, this is true for other genome-wide datasets from similar populations, e.g. the HapMap samples of Han Chinese and Japanese have been merged in all analyses due to their high genetic resemblance [Frazer et al., 2007]. Altogether, our analyses, fuelled by the results of genome-wide datasets, indicate that the genetic differentiation observed among subpopulations from the same continent is weak enough not to influence any of our conclusions regarding the action of natural selection.

1
2
3 1 The first interesting observation that can be made from this study is the absence of non-
4
5 2 synonymous mutations in *IFNG*. The strong selective constraint acting on *IFNG* is supported by
6
7 3 our inter-species analyses that clearly indicated that *IFNG* has been subject to intense purifying
8
9 4 selection. In addition, when considering a genome-wide resequencing dataset of genes showing
10
11 5 similar features than *IFNG* [Bustamante et al., 2005], this gene falls into the ~10% of genes
12
13 6 involved in immune system processes that display the strongest selective constraints on amino-
14
15 7 acid variation. Interestingly, the intensity of purifying selection on *IFNG* estimated from our
16
17 8 dataset ($\omega = 0.0189$) is extremely similar to that obtained using another population panel ($\omega =$
18
19 9 0.0184 from [Bustamante et al., 2005]), indicating that the detection of strong purifying selection
20
21 10 is not sensitive to the population considered. The fact that nonsynonymous mutations are not
22
23 11 tolerated suggests that amino-acid replacements at *IFNG* may have fatal consequences and be
24
25 12 quickly removed from the population. Such extreme protein conservation makes this gene an
26
27 13 excellent candidate to be involved in severe, rather lethal, diseases. A similar situation has been
28
29 14 observed for toll-like receptor 3 (TLR3); this gene is under purifying selection [Barreiro et al.,
30
31 15 2009] and TLR3 defects confer predisposition to childhood herpes simplex encephalitis [Zhang et
32
33 16 al., 2007]. Altogether, our data indicate that IFN- γ plays an essential, non-redundant role in host
34
35 17 survival. Because it has been proposed that IFN- γ plays a key role against mycobacterial
36
37 18 infections, but has a smaller impact on viral clearance [Dorman et al., 2004; Filipe-Santos et al.,
38
39 19 2006; Zhang et al., 2008], it is likely that IFN- γ is essential and non-redundant for protective
40
41 20 immunity against mycobacterial diseases.

42
43 21 Leaving aside the strong selective constraint maintaining intact the IFN- γ protein
44
45 22 sequence at the species-wide level, our analyses unmasked more subtle evolutionary events
46
47 23 acting at the population-specific level. For example, we identified an excess of high-frequency
48
49 24 derived alleles in *IFNG* in Africa, as attested by the significance of the Fay and Wu's *H* test using
50
51
52
53
54
55
56
57
58
59
60

1 the [Voight et al., 2005] model. This pattern was accounted for by the presence of 3 African
2 chromosomes each carrying 3 SNPs at the ancestral state, i.e. +3610A>G, +4802G>A, and
3 +5164C>T (c.[367-519A>G; *539G>A; *901C>T]), while the remaining chromosomes, in both
4 Africans and Eurasians, harbor the derived state at these three positions (Supp. Figure S5). Two
5 plausible explanations can be put forward to explain this pattern. First, it may testify the
6 occurrence of an almost-complete selective sweep worldwide, attesting for a selective advantage
7 associated with the derived states at the 3 SNPs. Although none of these SNPs corresponds to
8 mutations changing the amino-acid sequence, they could themselves have functional
9 consequences or be associated with mutations located further away in regulatory regions.
10 Alternatively, this pattern could also result from past population demographic events, such as
11 population structure within Africa [Wall and Hammer, 2006] and intercontinental migration
12 [Zeng et al., 2006]. Indeed, when using the model of [Laval et al., 2010], which assumes
13 intercontinental gene flow, the Fay and Wu's H at *IFNG* lost its significance in Africa (Table 1).
14 Because the differences in significance at *IFNG* could result from other parameter estimates that
15 differ between the two demographic models (e.g. intensity of the out-of-Africa bottleneck), we
16 estimated the P -value of Fay & Wu's H at *IFNG* using the model of [Laval et al., 2010] but
17 assuming no intercontinental migration. In this case, similarly to [Voight et al., 2005], we
18 obtained a significant Fay and Wu's H in Africa ($P \leq 0.05$), indicating that non-negligible
19 intercontinental migration could explain the patterns observed. Even if the ancestral alleles of
20 these 3 SNPs are absent from European and Asian individuals of our panel, the intercontinental
21 migration scenario is likely, since 2 of these 3 alleles do segregate in some European and/or
22 Asian individuals from the HapMap Phase III data (Supp. Table S4).

23 Our DIND analyses identified a population-specific signature of positive selection
24 targeting *IFNG* among Europeans, specifically at the SNPs +874T (c.115-483A>T) and +5173G

1 (c.*910A>G) (Figure 2B). We failed in detecting departures from neutrality based on classical
2 neutrality tests, most likely because of their inadequate power under a scenario of positive
3 selection acting on standing variation [Pritchard et al., 2010; Przeworski et al., 2005]. Indeed,
4 both SNPs +874T and +5173G were already present before the out-of-Africa exodus, as attested
5 by their appreciable frequencies among African populations, a situation that explains the low F_{ST}
6 values of these two SNPs between Europeans and the other populations. However, the fact that
7 the SNP +874A>T is well known to have functional consequences further reinforces our
8 population genetics prediction. Indeed, the +874T allele has been shown to provide a binding site
9 for the transcription factor NF- κ B, and to be associated with both higher production of IFN- γ
10 [Pravica et al., 2000] and higher resistance against intracellular pathogens such as
11 *Mycobacterium tuberculosis* [Ding et al., 2008; Etokebe et al., 2006; Lopez-Maderuelo et al.,
12 2003; Rossouw et al., 2003; Sallakci et al., 2007; Tso et al., 2005]. Further support to the action
13 of positive selection targeting the +874T allele comes from our survey of associated
14 microsatellite variation. Indeed, despite the chromosomes harboring the +874T allele are very
15 frequent in Europe (54%), they present a lower microsatellite diversity than those harboring the
16 +874A allele (expected heterozygosities in Europe of 0.17 vs. 0.26, respectively). Our results are
17 collectively consistent with a selective advantage for a higher production of IFN- γ in Europeans,
18 suggesting the existence of different, or stronger, selective pressures in Europe associated with
19 IFN- γ production. Taken together, although a strong selective constraint prevents qualitative
20 changes of the IFN- γ protein, there is a fine-tuned regulation of IFN- γ expression that seems to
21 evolve adaptively.

22 In contrast to *IFNG*, where nonsynonymous mutations are not tolerated, our results
23 unmasked more relaxed selective constraints at *IFNGR1* and *IFNGR2*, where we observed 10 and
24 6 nonsynonymous mutations, respectively (Table 2). However, several lines of evidence support

1 the notion that these two genes are not entirely free to accumulate functional variation. *IFNGR1*
2 appears to evolve under the action of weak negative selection, indicating that nonsynonymous
3 mutations, although tolerated, are kept at low population frequencies because they may have
4 weakly deleterious effects. In turn, amino-acid variation at *IFNGR2* seems to be somehow
5 constrained, as attested by the value of ω that is lower than 1. In addition, the only
6 nonsynonymous mutation observed at *IFNGR2* that is found at a high population frequency is
7 predicted to be benign (i.e., mutation likely not to impact protein function or weakly deleterious),
8 an observation that indicates that mutations leading to major changes in protein function are not
9 allowed to increase in frequency in the population. From the clinical genetics angle, several other
10 mutations in *IFNGR1* and *IFNGR2* have been shown to be associated with impaired cellular
11 responses to IFN- γ and to result in Mendelian susceptibility to mycobacterial disease (see [Zhang
12 et al., 2008] for an extensive review). The level of cellular responsiveness to IFN- γ seems to
13 strongly correlate with the clinical severity of mycobacterial disease; e.g., patients with complete
14 IFN- γ R1 or IFN- γ R2 deficiency display mycobacterial diseases early in life and have a poor
15 prognosis. Altogether, population and clinical data clearly show that no variation having a
16 significant impact on protein function is tolerated at *IFNGR1* and *IFNGR2*, highlighting more
17 generally that the entire IFN- γ pathway is essential in host survival.

18 Then, we have also observed population-specific signatures of positive selection in both
19 genes encoding the IFN- γ receptor. *IFNGR1* displays signatures of positive selection in Africa as
20 attested by the results of XP-EHH [Sabeti et al., 2007] (Supp. Figure S4). The observation that
21 the SNP +130G>A (c.85+45G>A) presents the highest levels of population differentiation
22 between African and non-African populations (e.g., $F_{ST}=0.45$, African versus Asian) suggests
23 that this SNP, or another in LD, could be the target of positive selection. If the SNP +130G>A
24 (c.85+45G>A) was the genuine selected allele, it is interesting to note that natural selection

1 would have increased the frequency of the ancestral allele [Di Rienzo, 2006] up to 95.2% in
2 contemporary African populations. We have also observed a population-specific signature of
3 positive selection in the gene encoding the second subunit of the IFN- γ receptor. Indeed, we
4 identified a strong signature of positive selection in Asian populations. This signature is most
5 likely explained by the intronic SNP +23133G>A (c.413-209G>A), as attested by the significant
6 values obtained for both the DIND test as well as the levels of population differentiation (Figs. 2
7 and 3). The derived allele at the SNP +23133 is absent in Africa but reaches 37% in Asia,
8 therefore indicating that this mutation likely appeared after the Out-of-Africa exodus. The
9 functional characterization of these two variants is now needed.

10 Taken together, the integration of our population genetics data into a clinical framework
11 clearly demonstrates that the IFN- γ pathway is essential and non-redundant in host defense, most
12 likely in protective immunity against mycobacteria. Future population genetics data will shed
13 light on how redundant or essential in host defense are the other multiple members of the human
14 IFN family, including type-I and type-III IFNs.

1
2
3 **1 ACKNOWLEDGEMENTS**
4

5
6 2 We thank Jean-Laurent Casanova, Eileen Hoal, Roberto Toro and Sandra Pellegrini for helpful
7
8 3 suggestions and for critical reading of the manuscript. This work has been supported by the
9
10 4 Institut Pasteur, the ANR (ANR-08-MIEN-009-01), the Fondation pour la Recherche Médicale,
11
12 5 the CNRS, Merck-Serono, and a EPFL-Debiopharm Life Sciences Award to L.Q.-M.
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

1
2
3 **1 REFERENCES**

- 4
5 2 Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS,
6
7
8 3 Sunyaev SR. 2010. A method and server for predicting damaging missense mutations.
9
10 4 Nat Methods 7(4):248-249.
- 11
12 5 Altshuler DM, Gibbs RA, Peltonen L, Dermitzakis E, Schaffner SF, Yu F, Bonnen PE, de Bakker
13
14
15 6 PI, Deloukas P, Gabriel SB, Gwilliam R, Hunt S, Inouye M, Jia X, Palotie A, Parkin M,
16
17 7 Whittaker P, Chang K, Hawes A, Lewis LR, Ren Y, Wheeler D, Muzny DM, Barnes C,
18
19 8 Darvishi K, Hurles M, Korn JM, Kristiansson K, Lee C, McCarroll SA, Nemesh J, Keinan
20
21 9 A, Montgomery SB, Pollack S, Price AL, Soranzo N, Gonzaga-Jauregui C, Anttila V,
22
23 10 Brodeur W, Daly MJ, Leslie S, McVean G, Moutsianas L, Nguyen H, Zhang Q, Ghorri
24
25 11 MJ, McGinnis R, McLaren W, Takeuchi F, Grossman SR, Shlyakhter I, Hostetter EB,
26
27 12 Sabeti PC, Adebamowo CA, Foster MW, Gordon DR, Licinio J, Manca MC, Marshall
28
29 13 PA, Matsuda I, Ngare D, Wang VO, Reddy D, Rotimi CN, Royal CD, Sharp RR, Zeng C,
30
31 14 Brooks LD, McEwen JE. 2010. Integrating common and rare genetic variation in diverse
32
33 15 human populations. Nature 467(7311):52-58.
- 34
35 16 Barreiro LB, Ben-Ali M, Quach H, Laval G, Patin E, Pickrell JK, Bouchier C, Tichit M,
36
37 17 Neyrolles O, Gicquel B, Kidd JR, Kidd KK, Alcais A, Ragimbeau J, Pellegrini S, Abel L,
38
39 18 Casanova JL, Quintana-Murci L. 2009. Evolutionary dynamics of human Toll-like
40
41 19 receptors and their different contributions to host defense. PLoS Genet 5(7):e1000562.
- 42
43 20 Barreiro LB, Laval G, Quach H, Patin E, Quintana-Murci L. 2008. Natural selection has driven
44
45 21 population differentiation in modern humans. Nat Genet 40(3):340-345.
- 46
47 22 Barrett JC, Fry B, Maller J, Daly MJ. 2005. Haploview: analysis and visualization of LD and
48
49 23 haplotype maps. Bioinformatics 21(2):263-265.
- 50
51
52
53
54
55
56
57
58
59
60

- 1 Bustamante CD, Fledel-Alon A, Williamson S, Nielsen R, Hubisz MT, Glanowski S, Tanenbaum
2 DM, White TJ, Sninsky JJ, Hernandez RD, Civello D, Adams MD, Cargill M, Clark AG.
3 2005. Natural selection on protein-coding genes in the human genome. *Nature*
4 437(7062):1153-1157.
- 5 Bustamante CD, Nielsen R, Sawyer SA, Olsen KM, Purugganan MD, Hartl DL. 2002. The cost
6 of inbreeding in *Arabidopsis*. *Nature* 416(6880):531-534.
- 7 Cann HM, de Toma C, Cazes L, Legrand MF, Morel V, Piouffre L, Bodmer J, Bodmer WF,
8 Bonne-Tamir B, Cambon-Thomsen A, Chen Z, Chu J, Carcassi C, Contu L, Du R,
9 Excoffier L, Ferrara GB, Friedlaender JS, Groot H, Gurwitz D, Jenkins T, Herrera RJ,
10 Huang X, Kidd J, Kidd KK, Langaney A, Lin AA, Mehdi SQ, Parham P, Piazza A,
11 Pistillo MP, Qian Y, Shu Q, Xu J, Zhu S, Weber JL, Greely HT, Feldman MW, Thomas
12 G, Dausset J, Cavalli-Sforza LL. 2002. A human genome diversity cell line panel. *Science*
13 296(5566):261-262.
- 14 den Dunnen JT, Antonarakis SE. 2000. Mutation nomenclature extensions and suggestions to
15 describe complex mutations: a discussion. *Hum Mutat* 15(1):7-12.
- 16 Di Rienzo A. 2006. Population genetics models of common diseases. *Curr Opin Genet Dev*
17 16(6):630-636.
- 18 Ding S, Li L, Zhu X. 2008. Polymorphism of the interferon-gamma gene and risk of tuberculosis
19 in a southeastern Chinese population. *Hum Immunol* 69(2):129-133.
- 20 Dorman SE, Picard C, Lammas D, Heyne K, van Dissel JT, Baretto R, Rosenzweig SD, Newport
21 M, Levin M, Roesler J, Kumararatne D, Casanova JL, Holland SM. 2004. Clinical
22 features of dominant and recessive interferon gamma receptor 1 deficiencies. *Lancet*
23 364(9451):2113-2121.

- 1
2
3 1 Durbin RM, Abecasis GR, Altshuler DL, Auton A, Brooks LD, Gibbs RA, Hurles ME, McVean
4
5 2 GA. 2010. A map of human genome variation from population-scale sequencing. *Nature*
6
7 467(7319):1061-1073.
8
9
10 4 Etokebe GE, Bulat-Kardum L, Johansen MS, Knezevic J, Balen S, Matakovic-Mileusnic N,
11
12 Matanic D, Flego V, Pavelic J, Beg-Zec Z, Dembic Z. 2006. Interferon-gamma gene
13
14 (T874A and G2109A) polymorphisms are associated with microscopy-positive
15
16 (T874A and G2109A) polymorphisms are associated with microscopy-positive
17
18 tuberculosis. *Scand J Immunol* 63(2):136-141.
19
20 8 Excoffier L, Laval G, Schneider S. 2005. Arlequin (version 3.0): An integrated software package
21
22 for population genetics data analysis. *Evol Bioinform Online* 1:47-50.
23
24 10 Excoffier L, Smouse PE, Quattro JM. 1992. Analysis of molecular variance inferred from metric
25
26 distances among DNA haplotypes: application to human mitochondrial DNA restriction
27
28 data. *Genetics* 131(2):479-491.
29
30
31 13 Filipe-Santos O, Bustamante J, Chagnier A, Vogt G, de Beaucoudrey L, Feinberg J, Jouanguy E,
32
33 Boisson-Dupuis S, Fieschi C, Picard C, Casanova JL. 2006. Inborn errors of IL-12/23-
34
35 and IFN-gamma-mediated immunity: molecular, cellular, and clinical features. *Semin*
36
37 *Immunol* 18(6):347-361.
38
39
40 17 Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A,
41
42 Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao
43
44 Y, Hu H, Hu W, Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B,
45
46 Zhang Q, Zhao H, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A,
47
48 Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M,
49
50 Roy J, Stahl E, Winchester E, Ziaugra L, Altshuler D, Shen Y, Yao Z, Huang W, Chu X,
51
52 He Y, Jin L, Liu Y, Shen Y, Sun W, Wang H, Wang Y, Wang Y, Xiong X, Xu L, Waye
53
54 MM, Tsui SK, Xue H, Wong JT, Galver LM, Fan JB, Gunderson K, Murray SS, Oliphant
55
56
57
58
59
60

1
2
3 1 AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier JF, Phillips MS,
4
5 2 Roumy S, Sallee C, Verner A, Hudson TJ, Kwok PY, Cai D, Koboldt DC, Miller RD,
6
7 3 Pawlikowska L, Taillon-Miller P, Xiao M, Tsui LC, Mak W, Song YQ, Tam PK,
8
9 4 Nakamura Y, Kawaguchi T, Kitamoto T, Morizono T, Nagashima A, Ohnishi Y, Sekine
10
11 5 A, Tanaka T, Tsunoda T, Deloukas P, Bird CP, Delgado M, Dermitzakis ET, Gwilliam R,
12
13 6 Hunt S, Morrison J, Powell D, Stranger BE, Whittaker P, Bentley DR, Daly MJ, de
14
15 7 Bakker PI, Barrett J, Chretien YR, Maller J, McCarroll S, Patterson N, Pe'er I, Price A,
16
17 8 Purcell S, Richter DJ, Sabeti P, Saxena R, Schaffner SF, Sham PC, Varilly P, Altshuler D,
18
19 9 Stein LD, Krishnan L, Smith AV, Tello-Ruiz MK, Thorisson GA, Chakravarti A, Chen
20
21 10 PE, Cutler DJ, Kashuk CS, Lin S, Abecasis GR, Guan W, Li Y, Munro HM, Qin ZS,
22
23 11 Thomas DJ, McVean G, Auton A, Bottolo L, Cardin N, Eyheramendy S, Freeman C,
24
25 12 Marchini J, Myers S, Spencer C, Stephens M, Donnelly P, Cardon LR, Clarke G, Evans
26
27 13 DM, Morris AP, Weir BS, Tsunoda T, Mullikin JC, Sherry ST, Feolo M, Skol A, Zhang
28
29 14 H, Zeng C, Zhao H, Matsuda I, Fukushima Y, Macer DR, Suda E, Rotimi CN,
30
31 15 Adebamowo CA, Ajayi I, Aniagwu T, Marshall PA, Nkwodimmah C, Royal CD, Leppert
32
33 16 MF, Dixon M, Peiffer A, Qiu R, Kent A, Kato K, Niikawa N, Adewole IF, Knoppers BM,
34
35 17 Foster MW, Clayton EW, Watkin J, Gibbs RA, Belmont JW, Muzny D, Nazareth L,
36
37 18 Sodergren E, Weinstock GM, Wheeler DA, Yakub I, Gabriel SB, Onofrio RC, Richter
38
39 19 DJ, Ziaugra L, Birren BW, Daly MJ, Altshuler D, Wilson RK, Fulton LL, Rogers J,
40
41 20 Burton J, Carter NP, Clee CM, Griffiths M, Jones MC, McLay K, Plumb RW, Ross MT,
42
43 21 Sims SK, Willey DL, Chen Z, Han H, Kang L, Godbout M, Wallenburg JC,
44
45 22 L'Archeveque P, Bellemare G, Saeki K, Wang H, An D, Fu H, Li Q, Wang Z, Wang R,
46
47 23 Holden AL, Brooks LD, McEwen JE, Guyer MS, Wang VO, Peterson JL, Shi M, Spiegel
48
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3 1 J, Sung LM, Zacharia LF, Collins FS, Kennedy K, Jamieson R, Stewart J. 2007. A second
4
5 2 generation human haplotype map of over 3.1 million SNPs. *Nature* 449(7164):851-861.
6
7
8 3 Gilad Y, Bustamante CD, Lancet D, Paabo S. 2003. Natural selection on the olfactory receptor
9
10 4 gene family in humans and chimpanzees. *Am J Hum Genet* 73(3):489-501.
11
12 5 Griffiths RC, Tavaré S. 1994. Sampling theory for neutral alleles in a varying environment.
13
14 6 *Philos Trans R Soc Lond B Biol Sci* 344(1310):403-410.
15
16 7 International-HapMap-Consortium. 2005. A haplotype map of the human genome. *Nature*
17
18 8 437(7063):1299-1320.
19
20 9 Key LL, Jr., Ries WL, Rodriguiz RM, Hatcher HC. 1992. Recombinant human interferon gamma
21
22 10 therapy for osteopetrosis. *J Pediatr* 121(1):119-124.
23
24
25 11 Kreitman M. 2000. Methods to detect selection in populations with applications to the human.
26
27 12 *Annu Rev Genomics Hum Genet* 1:539-559.
28
29 13 Laval G, Excoffier L. 2004. SIMCOAL 2.0: a program to simulate genomic diversity over large
30
31 14 recombining regions in a subdivided population with a complex history. *Bioinformatics*
32
33 15 20(15):2485-2487.
34
35 16 Laval G, Patin E, Barreiro LB, Quintana-Murci L. 2010. Formulating a historical and
36
37 17 demographic model of recent human evolution based on resequencing data from
38
39 18 noncoding regions. *PLoS One* 5(4):e10284.
40
41 19 Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, Cann HM, Barsh GS,
42
43 20 Feldman M, Cavalli-Sforza LL, Myers RM. 2008. Worldwide human relationships
44
45 21 inferred from genome-wide patterns of variation. *Science* 319(5866):1100-1104.
46
47
48 22 Lopez-Maderuelo D, Arnalich F, Serantes R, Gonzalez A, Codoceo R, Madero R, Vazquez JJ,
49
50 23 Montiel C. 2003. Interferon-gamma and interleukin-10 gene polymorphisms in
51
52 24 pulmonary tuberculosis. *Am J Respir Crit Care Med* 167(7):970-975.
53
54
55
56
57
58
59
60

- 1
2
3 1 Myers S, Bottolo L, Freeman C, McVean G, Donnelly P. 2005. A fine-scale map of
4
5 2 recombination rates and hotspots across the human genome. *Science* 310(5746):321-324.
6
7
8 3 Nielsen R, Hellmann I, Hubisz M, Bustamante C, Clark AG. 2007. Recent and ongoing selection
9
10 4 in the human genome. *Nat Rev Genet* 8(11):857-868.
11
12
13 5 Pestka S, Krause CD, Walter MR. 2004. Interferons, interferon-like cytokines, and their
14
15 6 receptors. *Immunol Rev* 202:8-32.
16
17
18 7 Pravica V, Asderakis A, Perrey C, Hajeer A, Sinnott PJ, Hutchinson IV. 1999. In vitro production
19
20 8 of IFN-gamma correlates with CA repeat polymorphism in the human IFN-gamma gene.
21
22 9 *Eur J Immunogenet* 26(1):1-3.
23
24
25 10 Pravica V, Perrey C, Stevens A, Lee JH, Hutchinson IV. 2000. A single nucleotide
26
27 11 polymorphism in the first intron of the human IFN-gamma gene: absolute correlation with
28
29 12 a polymorphic CA microsatellite marker of high IFN-gamma production. *Hum Immunol*
30
31 13 61(9):863-866.
32
33
34 14 Pritchard JK, Pickrell JK, Coop G. 2010. The genetics of human adaptation: hard sweeps, soft
35
36 15 sweeps, and polygenic adaptation. *Curr Biol* 20(4):R208-215.
37
38
39 16 Przeworski M, Coop G, Wall JD. 2005. The signature of positive selection on standing genetic
40
41 17 variation. *Evolution* 59(11):2312-2323.
42
43
44 18 Rossouw M, Nel HJ, Cooke GS, van Helden PD, Hoal EG. 2003. Association between
45
46 19 tuberculosis and a polymorphic NFkappaB binding site in the interferon gamma gene.
47
48 20 *Lancet* 361(9372):1871-1872.
49
50
51 21 Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R. 2003. DnaSP, DNA polymorphism
52
53 22 analyses by the coalescent and other methods. *Bioinformatics* 19(18):2496-2497.
54
55
56 23 Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, Gabriel SB, Platko JV,
57
58 24 Patterson NJ, McDonald GJ, Ackerman HC, Campbell SJ, Altshuler D, Cooper R,

- 1
2
3 1 Kwiatkowski D, Ward R, Lander ES. 2002. Detecting recent positive selection in the
4
5 2 human genome from haplotype structure. *Nature* 419(6909):832-837.
6
7
8 3 Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilly P, Shamovsky O, Palma A, Mikkelsen TS,
9
10 4 Altshuler D, Lander ES. 2006. Positive natural selection in the human lineage. *Science*
11
12 5 312(5780):1614-1620.
13
14
15 6 Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie X, Byrne EH, McCarroll
16
17 7 SA, Gaudet R, Schaffner SF, Lander ES, Frazer KA, Ballinger DG, Cox DR, Hinds DA,
18
19 8 Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S,
20
21 9 Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu
22
23 10 S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang Q, Zhao H, Zhao H, Zhou J,
24
25 11 Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M,
26
27 12 Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E,
28
29 13 Ziaugra L, Altshuler D, Shen Y, Yao Z, Huang W, Chu X, He Y, Jin L, Liu Y, Shen Y,
30
31 14 Sun W, Wang H, Wang Y, Wang Y, Xiong X, Xu L, Wayne MM, Tsui SK, Xue H, Wong
32
33 15 JT, Galver LM, Fan JB, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A,
34
35 16 Chagnon F, Ferretti V, Leboeuf M, Olivier JF, Phillips MS, Roumy S, Sallee C, Verner
36
37 17 A, Hudson TJ, Kwok PY, Cai D, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller
38
39 18 P, Xiao M, Tsui LC, Mak W, Song YQ, Tam PK, Nakamura Y, Kawaguchi T, Kitamoto
40
41 19 T, Morizono T, Nagashima A, Ohnishi Y, Sekine A, Tanaka T, Tsunoda T, Deloukas P,
42
43 20 Bird CP, Delgado M, Dermitzakis ET, Gwilliam R, Hunt S, Morrison J, Powell D,
44
45 21 Stranger BE, Whittaker P, Bentley DR, Daly MJ, de Bakker PI, Barrett J, Chretien YR,
46
47 22 Maller J, McCarroll S, Patterson N, Pe'er I, Price A, Purcell S, Richter DJ, Sabeti P,
48
49 23 Saxena R, Schaffner SF, Sham PC, Varilly P, Altshuler D, Stein LD, Krishnan L, Smith
50
51 24 AV, Tello-Ruiz MK, Thorisson GA, Chakravarti A, Chen PE, Cutler DJ, Kashuk CS, Lin
52
53
54
55
56
57
58
59
60

- 1
2
3 1 S, Abecasis GR, Guan W, Li Y, Munro HM, Qin ZS, Thomas DJ, McVean G, Auton A,
4
5 2 Bottolo L, Cardin N, Eyheramendy S, Freeman C, Marchini J, Myers S, Spencer C,
6
7 3 Stephens M, Donnelly P, Cardon LR, Clarke G, Evans DM, Morris AP, Weir BS,
8
9 4 Tsunoda T, Johnson TA, Mullikin JC, Sherry ST, Feolo M, Skol A, Zhang H, Zeng C,
10
11 5 Zhao H, Matsuda I, Fukushima Y, Macer DR, Suda E, Rotimi CN, Adebamowo CA,
12
13 6 Ajayi I, Aniagwu T, Marshall PA, Nkwodimmah C, Royal CD, Leppert MF, Dixon M,
14
15 7 Peiffer A, Qiu R, Kent A, Kato K, Niikawa N, Adewole IF, Knoppers BM, Foster MW,
16
17 8 Clayton EW, Watkin J, Gibbs RA, Belmont JW, Muzny D, Nazareth L, Sodergren E,
18
19 9 Weinstock GM, Wheeler DA, Yakub I, Gabriel SB, Onofrio RC, Richter DJ, Ziaugra L,
20
21 10 Birren BW, Daly MJ, Altshuler D, Wilson RK, Fulton LL, Rogers J, Burton J, Carter NP,
22
23 11 Clee CM, Griffiths M, Jones MC, McLay K, Plumb RW, Ross MT, Sims SK, Willey DL,
24
25 12 Chen Z, Han H, Kang L, Godbout M, Wallenburg JC, L'Archeveque P, Bellemare G,
26
27 13 Saeki K, Wang H, An D, Fu H, Li Q, Wang Z, Wang R, Holden AL, Brooks LD,
28
29 14 McEwen JE, Guyer MS, Wang VO, Peterson JL, Shi M, Spiegel J, Sung LM, Zacharia
30
31 15 LF, Collins FS, Kennedy K, Jamieson R, Stewart J. 2007. Genome-wide detection and
32
33 16 characterization of positive selection in human populations. *Nature* 449(7164):913-918.
34
35 17 Sallakci N, Coskun M, Berber Z, Gurkan F, Kocamaz H, Uysal G, Bhujju S, Yavuzer U, Singh M,
36
37 18 Yegin O. 2007. Interferon-gamma gene+874T-A polymorphism is associated with
38
39 19 tuberculosis and gamma interferon response. *Tuberculosis (Edinb)* 87(3):225-230.
40
41 20 Sawyer SA, Hartl DL. 1992. Population genetics of polymorphism and divergence. *Genetics*
42
43 21 132(4):1161-1176.
44
45 22 Schroder K, Hertzog PJ, Ravasi T, Hume DA. 2004. Interferon-gamma: an overview of signals,
46
47 23 mechanisms and functions. *J Leukoc Biol* 75(2):163-189.
48
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3 1 Stephens M, Donnelly P. 2003. A comparison of bayesian methods for haplotype reconstruction
4
5 2 from population genotype data. *Am J Hum Genet* 73(5):1162-1169.
6
7
8 3 Takahashi M, Matsuda F, Margetic N, Lathrop M. 2003. Automated identification of single
9
10 4 nucleotide polymorphisms from sequencing data. *J Bioinform Comput Biol* 1(2):253-265.
11
12
13 5 Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, Diemer K, Muruganujan
14
15 6 A, Narechania A. 2003. PANTHER: a library of protein families and subfamilies indexed
16
17 7 by function. *Genome Res* 13(9):2129-2141.
18
19
20 8 Todd PA, Goa KL. 1992. Interferon gamma-1b. A review of its pharmacology and therapeutic
21
22 9 potential in chronic granulomatous disease. *Drugs* 43(1):111-122.
23
24
25 10 Tso HW, Ip WK, Chong WP, Tam CM, Chiang AK, Lau YL. 2005. Association of interferon
26
27 11 gamma and interleukin 10 genes with tuberculosis in Hong Kong Chinese. *Genes Immun*
28
29 12 6(4):358-363.
30
31
32 13 Voight BF, Adams AM, Frisse LA, Qian Y, Hudson RR, Di Rienzo A. 2005. Interrogating
33
34 14 multiple aspects of variation in a full resequencing data set to infer human population size
35
36 15 changes. *Proc Natl Acad Sci U S A* 102(51):18508-18513.
37
38
39 16 Wall JD. 1999. Recombination and the power of statistical tests of neutrality. *Genetical Research*
40
41 17 74(1):65-79.
42
43
44 18 Wall JD, Hammer MF. 2006. Archaic admixture in the human genome. *Curr Opin Genet Dev*
45
46 19 16(6):606-610.
47
48
49 20 Weir BS. 1984. Estimating F-statistics for the analysis of population structure. *Evolution*
50
51 21 38(6):1358-1370.
52
53
54 22 Williamson SH, Hernandez R, Fledel-Alon A, Zhu L, Nielsen R, Bustamante CD. 2005.
55
56 23 Simultaneous inference of selection and population growth from patterns of variation in
57
58 24 the human genome. *Proc Natl Acad Sci U S A* 102(22):7882-7887.
59
60

- 1
2
3 1 Zeng K, Fu YX, Shi S, Wu CI. 2006. Statistical tests for detecting positive selection by utilizing
4
5
6 2 high-frequency variants. *Genetics* 174(3):1431-1439.
7
8 3 Zhang SY, Boisson-Dupuis S, Chagnier A, Yang K, Bustamante J, Puel A, Picard C, Abel L,
9
10 4 Jouanguy E, Casanova JL. 2008. Inborn errors of interferon (IFN)-mediated immunity in
11
12 5 humans: insights into the respective roles of IFN-alpha/beta, IFN-gamma, and IFN-
13
14 6 lambda in host defense. *Immunol Rev* 226:29-40.
15
16
17 7 Zhang SY, Jouanguy E, Ugolini S, Smahi A, Elain G, Romero P, Segal D, Sancho-Shimizu V,
18
19 8 Lorenzo L, Puel A, Picard C, Chagnier A, Plancoulaine S, Titeux M, Cognet C, von
20
21 9 Bernuth H, Ku CL, Casrouge A, Zhang XX, Barreiro L, Leonard J, Hamilton C, Lebon P,
22
23 10 Heron B, Vallee L, Quintana-Murci L, Hovnanian A, Rozenberg F, Vivier E, Geissmann
24
25 11 F, Tardieu M, Abel L, Casanova JL. 2007. TLR3 deficiency in patients with herpes
26
27 12 simplex encephalitis. *Science* 317(5844):1522-1527.
28
29
30
31
32 13
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

FIGURE LEGENDS**Figure 1. Estimation of the intensity of natural selection acting on *IFNG*, *IFNGR1* and**

***IFNGR2*.** (A) Strength of **interspecies** purifying selection, as measured by estimated ω values.

(B) Strength of **intraspecies** negative selection, as measured by the population selection

coefficient γ . Bars indicate 95% confidence intervals, and filled circles indicate genes with ω and

γ estimates significantly lower than 1 and 0, respectively.

Figure 2. Signature of positive selection at *IFNG*, *IFNGR1*, and *IFNGR2*. DIND test in

Africans (A,D,G), Europeans (B,E,H) and East-Asians (C,F,I). We plotted the $i\pi_A/i\pi_D$ values

against the Derived Allele Frequencies (DAFs). *P*-values were obtained by comparing the $i\pi_A/i\pi_D$

values for the 3 genes against the expected $i\pi_A/i\pi_D$ values obtained from 10^4 simulations

considering a previously validated demographic model [Voight et al., 2005]. The higher dashed

line of each graph corresponds to the 99th percentile, and the lower to the 95th percentile.

Figure 3. Levels of population differentiation, measured by F_{ST} . Population pair-wise

comparisons of *IFNG*, *IFNGR1* and *IFNGR2* SNPs for (A) Africans vs. Europeans, (B) Africans

vs. Asians and (C) Europeans vs. Asians. F_{ST} values are plotted against expected heterozygosity.

The dashed lines represent the 95th and 99th percentiles of the HGDP-CEPH genotyping dataset

using the same individuals (represented by the density area in blue) [Li et al., 2008]. Black dots

correspond to silent polymorphisms and red dots correspond to nonsynonymous polymorphisms.

Table 1. Mean diversity indices and neutrality tests across *IFNG*, *IFNGR1* and *IFNGR2* genomic regions

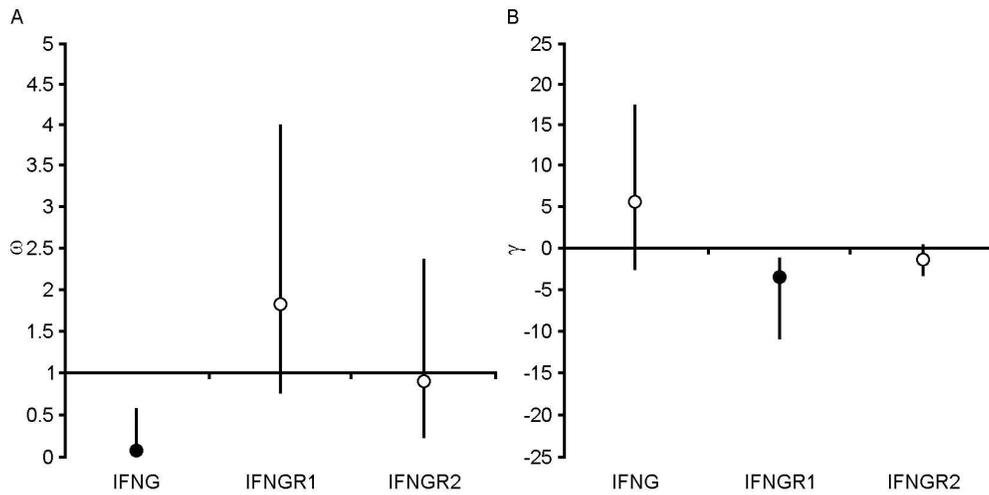
| | <i>IFNG</i> | | | | <i>IFNGR1</i> | | | | <i>IFNGR2</i> | | | |
|---|-------------------|-------------------|-----------------|-------------------|-------------------|-------------------|-----------------|-------------------|-------------------|-------------------|-----------------|-------------------|
| | Africa (N=124) | Europe (N=124) | Asia (N=124) | Global (N=372) | Africa (N=124) | Europe (N=124) | Asia (N=124) | Global (N=372) | Africa (N=124) | Europe (N=124) | Asia (N=124) | Global (N=372) |
| H | 17 | 8 | 7 | 26 | 25 | 12 | 16 | 43 | 25 | 17 | 18 | 49 |
| Hd | 0.71 | 0.64 | 0.64 | 0.72 | 0.82 | 0.69 | 0.81 | 0.81 | 0.90 | 0.79 | 0.74 | 0.87 |
| Syn | 0 | 0 | 0 | 0 | 1 | 1 | 2 | 2 | 4 | 0 | 3 | 7 |
| Non-syn | 0 | 0 | 0 | 0 | 8 | 1 | 3 | 10 | 5 | 3 | 1 | 6 |
| S | 21 | 9 | 6 | 28 | 25 | 11 | 17 | 42 | 29 | 16 | 19 | 47 |
| Singletons | 7 | 3 | 1 | 11 | 10 | 5 | 5 | 17 | 14 | 6 | 12 | 28 |
| INDELS | 2 | 1 | 1 | 2 | 3 | 1 | 1 | 5 | 4 | 1 | 1 | 4 |
| π (10^{-4}) | 3.7 | 4.2 | 3.2 | 4 | 4.3 | 4.1 | 5 | 4.7 | 6.9 | 7.1 | 5.5 | 7.2 |
| θ_w (10^{-4}) | 8.5 | 3.6 | 2.4 | 9.4 | 8.5 | 3.7 | 5.8 | 11.8 | 11.3 | 6.2 | 7.4 | 15.1 |
| TD | -1.60 | 0.34 | 0.70 | | -1.44 | 0.27 | -0.37 | | -1.13 | 0.39 | -0.73 | |
| D | -1.39 | -1.02 | 0.12 | | -2.13 | -2.02 | -0.95 | | -3.09 | -1.63 | -4.08**/†† | |
| F | -1.76 | -0.65 | 0.37 | | -2.23 | -1.43 | -0.88 | | -2.77 | -1.05 | -3.36**/†† | |
| H | -5.76* | -0.42 | -0.10 | | -0.82 | 0.71 | 1.26 | | -1.27 | -1.18 | -2.33 | |

N, number of chromosomes sequenced in the corresponding population; H, number of haplotypes; Hd, haplotype diversity; Syn, number of synonymous mutations; Non-syn, number of nonsynonymous mutations; S, number of segregating sites; INDELS, number of INDELS including the *IFNG* +875(CA)_n microsatellite; π , nucleotide diversity per site from average pairwise differences; θ_w , nucleotide diversity per site from number of segregating sites; TD, Tajima's *D*; D, Fu & Li's *D**; F, Fu & Li's *F**; H, Fay & Wu's *H*. **/**P*-values ≤ 0.01 and ≤ 0.05 , respectively, according to the model of [Voight, et al., 2005]; ††/† *P*-values ≤ 0.01 and ≤ 0.05 , respectively, according to the model of [Laval, et al., 2010]. *P*-values were obtained from coalescent simulations, according to the models proposed by [Voight, et al., 2005], which considers each continental population separately, and [Laval, et al., 2010], which considers inter-continental population migration.

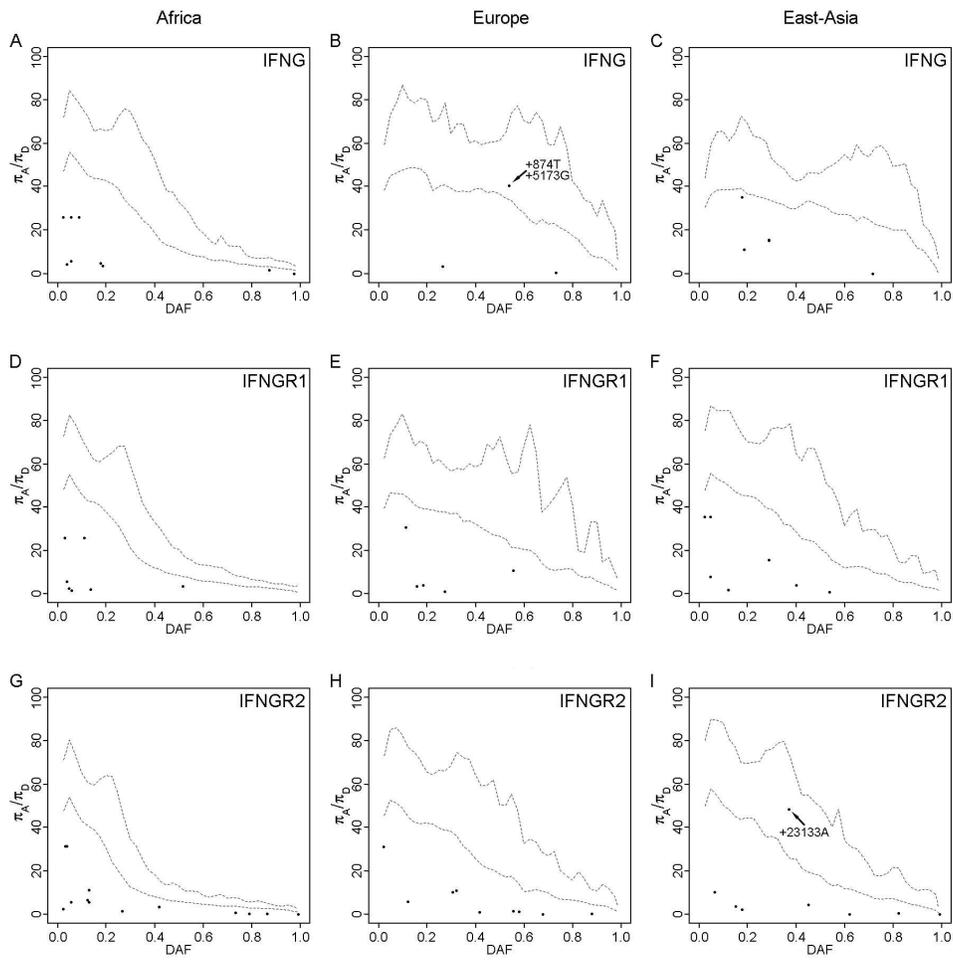
Table 2. Nonsynonymous changes in the *IFNGR1* and *IFNGR2* genes identified in this study.

| Gene | c.SNP | ATG position | Aminoacid change | Chromosomal location | Gene Location | Protein domain | Polyphen | Panther | dbSNP | Africa | Europe | Asia |
|---------------|-----------|--------------|------------------|----------------------|---------------|----------------|-------------|---------|------------|--------|--------|------|
| <i>IFNGR1</i> | c.5C>T | +5 | p.Ala2Val | chr6:137,540,460 | exon 1 | signal | po damaging | NA | | | | 0.8 |
| <i>IFNGR1</i> | c.40G>A | +40 | p.Val14Met | chr6:137,540,425 | exon 1 | signal | po damaging | 0.55613 | rs11575936 | | | 1.6 |
| <i>IFNGR1</i> | c.181G>A | +12346 | p.Val61Ile | chr6:137,528,119 | exon 2 | extracellular | pr damaging | 0.81834 | rs17175322 | 0.8 | | |
| <i>IFNGR1</i> | c.538G>A | +14988 | p.Gly180Arg | chr6:137,525,477 | exon 4 | extracellular | benign | 0.67155 | | 1.6 | | |
| <i>IFNGR1</i> | c.864C>G | +20691 | p.Ile288Met | chr6:137,519,774 | exon 7 | cytoplasmic | benign | 0.77695 | | 1.6 | | |
| <i>IFNGR1</i> | c.1004A>C | +20831 | p.His335Pro | chr6:137,519,634 | exon 7 | cytoplasmic | benign | 0.69608 | rs17175350 | 1.6 | | |
| <i>IFNGR1</i> | c.1027G>A | +20854 | p.Val343Met | chr6:137,519,611 | exon 7 | cytoplasmic | benign | 0.4295 | | 0.8 | | |
| <i>IFNGR1</i> | c.1034A>G | +20861 | p.His345Arg | chr6:137,519,604 | exon 7 | cytoplasmic | pr damaging | 0.48851 | | 0.8 | | |
| <i>IFNGR1</i> | c.1268G>A | +21095 | p.Ser423Asn | chr6:137,519,370 | exon 7 | cytoplasmic | benign | 0.85071 | | 0.8 | 0.8 | |
| <i>IFNGR1</i> | c.1400T>C | +21227 | p.Leu467Pro | chr6:137,519,238 | exon 7 | cytoplasmic | benign | NA | rs1887415 | 0.8 | | 4.8 |
| <i>IFNGR2</i> | c.173C>G | +11445 | p.Thr58Arg | chr21:34,787,294 | exon 3 | extracellular | po damaging | 0.49217 | rs4986958 | 12.9 | 0.8 | |
| <i>IFNGR2</i> | c.191G>A | +11463 | p.Arg64Gln | chr21:34,787,312 | exon 3 | extracellular | benign | 0.68807 | rs9808753 | 79.0 | 87.9 | 62.1 |
| <i>IFNGR2</i> | c.466A>C | +23395 | p.Ile156Leu | chr21:34,799,244 | exon 5 | extracellular | benign | 0.57445 | | | 1.6 | |
| <i>IFNGR2</i> | c.544A>G | +23473 | p.Lys182Glu | chr21:34,799,322 | exon 5 | extracellular | benign | 0.5363 | rs17878711 | 4.0 | | |
| <i>IFNGR2</i> | c.708A>T | +28781 | p.Glu236Asp | chr21:34,804,630 | exon 6 | transmembrane | po damaging | 0.69961 | | 0.8 | | |
| <i>IFNGR2</i> | c.889G>A | +33295 | p.Asp297Asn | chr21:34,809,144 | exon 8 | cytoplasmic | pr damaging | 0.76231 | | 0.8 | | |

A full description of all SNPs identified in this study (coding and non-coding) is available in Supp. Table S3. The position of each SNP was determined using the reference sequence listed in Supp. Table S2. The first amino-acid corresponds to the ancestral state, as defined considering the sequences of human, chimpanzee, gorilla, orangutan and rhesus. Chromosome location of each SNP is given according to the hg19 (GRCh37) human assembly. Protein domains are given by the UniProt Database. For the *IFNGR2* SNPs c.708A>T and c.889G>A, the domain is predicted by the UniProt Database as a potential domain. For the Polyphen analyses, “pr” stands for probably and “po” for possibly. For the Panther analyses, the P-deleterious values are shown. The frequencies, given in %, of each SNP in the different continental populations refer to the Derived Allele Frequency.



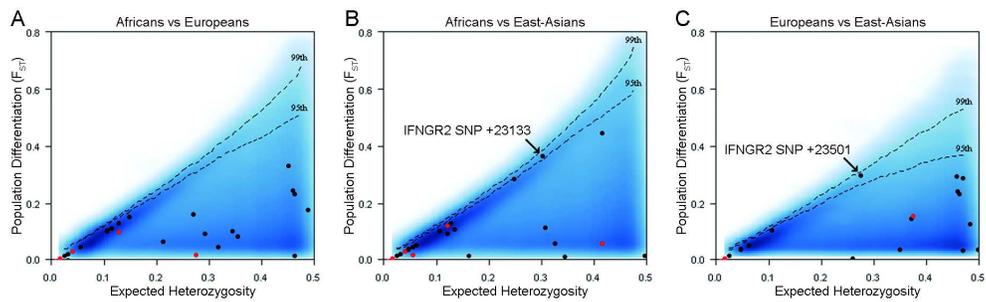
Estimation of the intensity of natural selection acting on IFNG, IFNGR1 and IFNGR2.
140x70mm (300 x 300 DPI)



Signature of positive selection at IFNG, IFNGR1, and IFNGR2.
160x160mm (300 x 300 DPI)



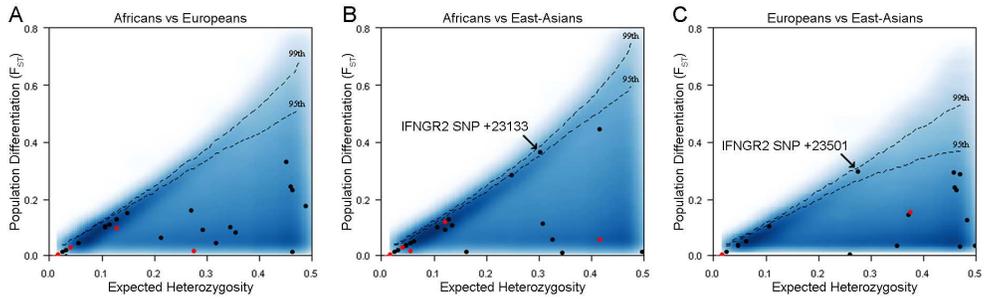
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



Levels of population differentiation, measured by F_{ST} .
160x53mm (300 x 300 DPI)

or Peer Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



Levels of population differentiation, measured by F_{ST} .
160x53mm (300 x 300 DPI)

or Peer Review

SUPPORTING MATERIAL FOR

Evolutionary genetics evidence of an essential, non-redundant role of the IFN- γ pathway in protective immunity

Jérémy Manry ^{1,2}, Guillaume Laval ^{1,2}, Etienne Patin ^{1,2}, Simona Fornarino ^{1,2}, Magali Tichit ³,
Christiane Bouchier ³, Luis B. Barreiro ⁴, Lluís Quintana-Murci ^{1,2}

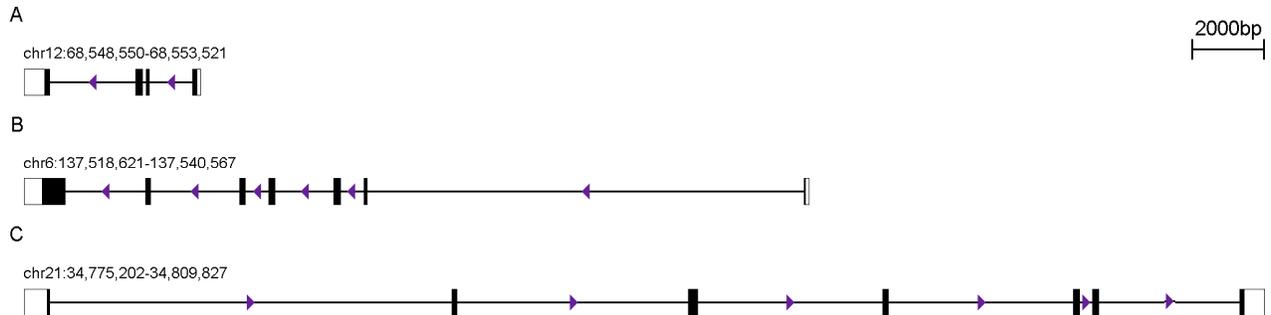
¹ Institut Pasteur, Human Evolutionary Genetics, Department of Genomes and Genetics, F-75015 Paris, France; ² Centre National de la Recherche Scientifique, URA3012, F-75015 Paris, France ; ³ Institut Pasteur, Plate-forme Génomique, Pasteur Genopole, Paris, France ; ⁴ Department of Human Genetics, University of Chicago, Chicago, USA

*Correspondence to Dr. Lluís Quintana-Murci, CNRS URA3012, UP Génétique Evolutive Humaine, Institut Pasteur, 25 rue du Dr. Roux, 75724 Paris Cedex 15, France ; Phone : +33.1.40.61.34.43 ; Fax : +33.1.45.68.86.39 ; E-mail : quintana@pasteur.fr

Short Title: Natural Selection acting on IFN- γ pathway

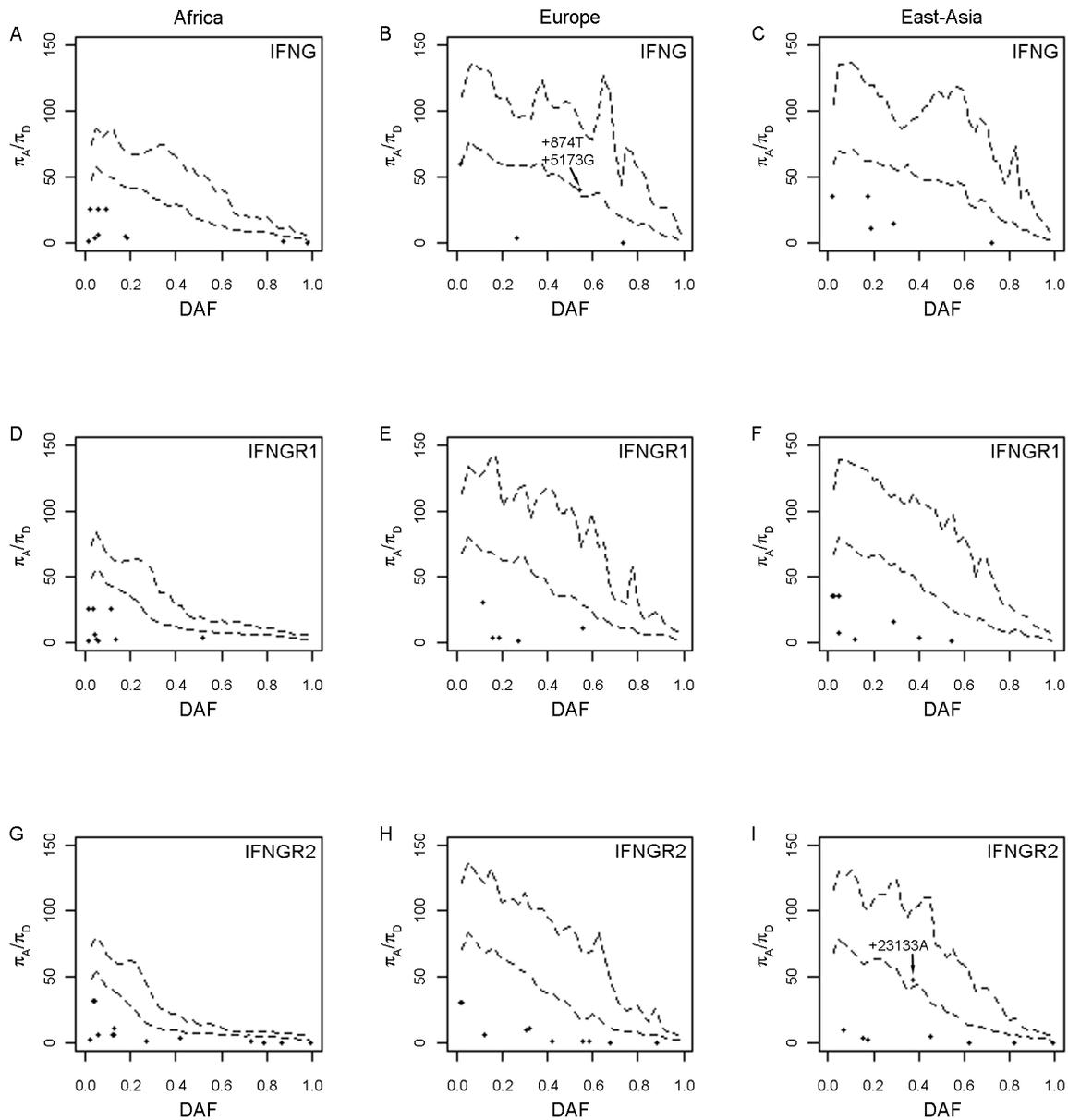
1
2
3 **Supp. Figure S1. Organization and structure of the *IFNG* (A), *IFNGR1* (B) and *IFNGR2***

4
5
6 **(C) genes**



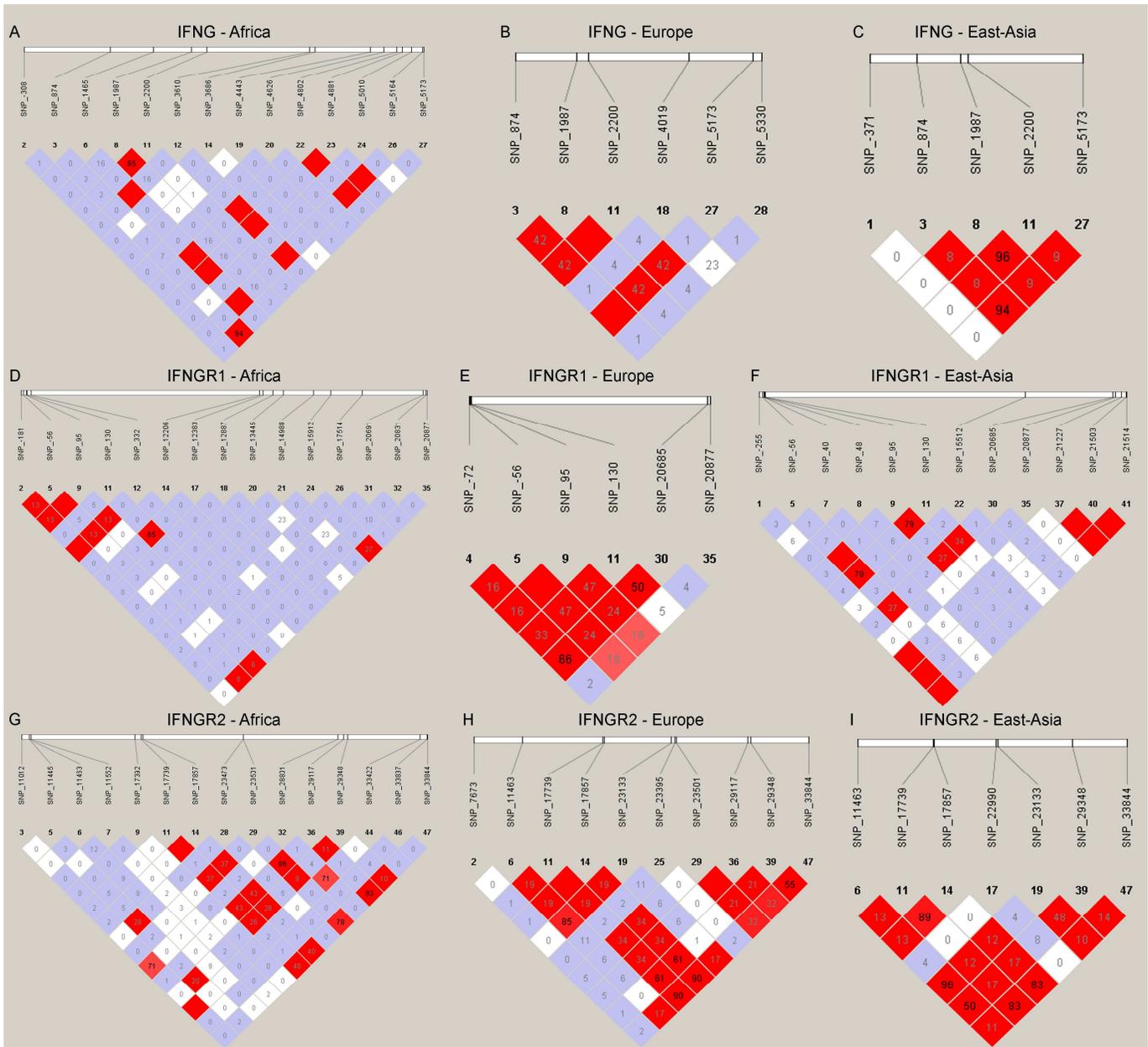
Black boxes correspond to coding exons, empty boxes to non-coding exons. Introns are represented by a line. The violet arrows indicate the orientation of the genes.

Supp. Figure S2. Signatures of positive selection, based on the DIND test, at *IFNG*, *IFNGR1*, and *IFNGR2* when considering the demographic model of [Laval et al., 2010]



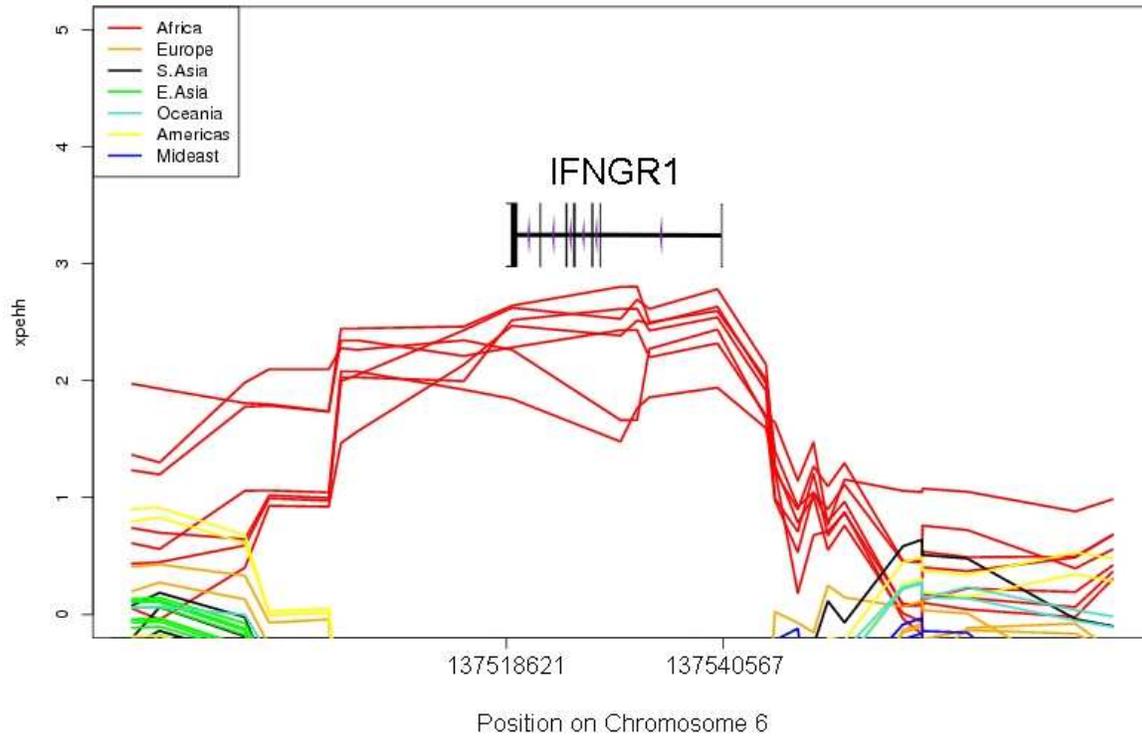
DIND test in Africans (A,D,G), Europeans (B,E,H) and East-Asians (C,F,I). We plotted the $i\pi_A/i\pi_D$ values against the Derived Allele Frequencies (DAFs). P -values were obtained by comparing the $i\pi_A/i\pi_D$ values for the 3 genes against the expected $i\pi_A/i\pi_D$ values obtained from 10^4 simulations considering a previously validated demographic model [Laval et al., 2010]. The higher dashed line of each graph corresponds to the 99th percentile, and the lower to the 95th percentile.

Supp. Figure S3. Linkage disequilibrium maps for the *IFNG* (A-C), *IFNGR1* (D-F) and *IFNGR2* (G-I) genes.



LD map in African (A, D, G), European (B, E, H), and East-Asian (C, F, I) populations. LD was estimated for SNPs with MAF>0.01. In each square, r^2 values are presented. Red squares without any value correspond to $r^2=1$.

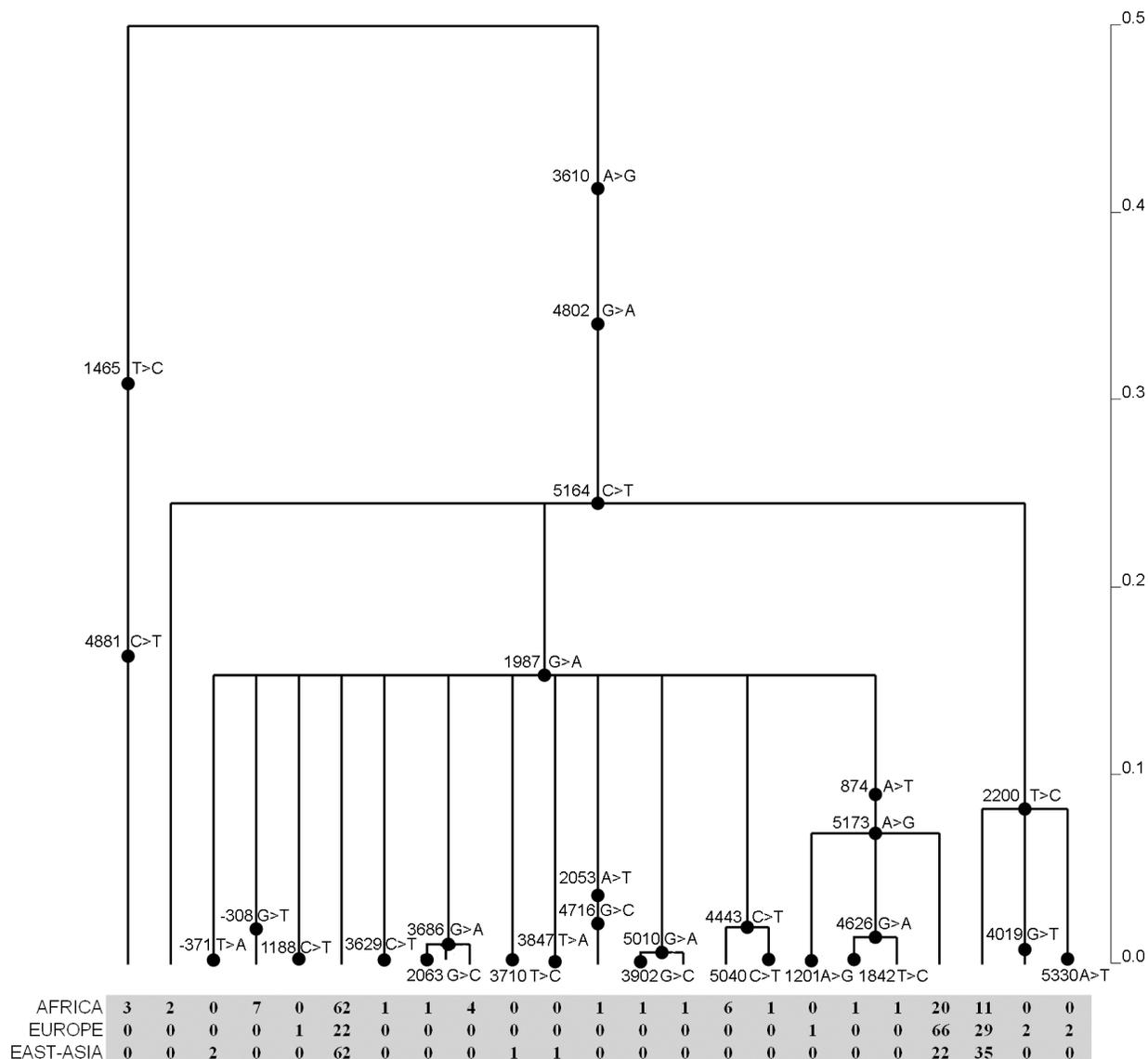
1
2
3
4 **Supp. Figure S4: Detection of positive selection acting on *IFNGR1* in Africa, using the XP-**
5
6 **EHH test.**
7
8
9
10



The region is centered on *IFNGR1*. This graph was obtained using the HGDP selection browser (<http://hgdp.uchicago.edu/cgi-bin/gbrowse/HGDP/>).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Supp. Figure S5. *IFNG* gene tree



Time is scaled in millions of years. Mutations are named for their physical positions along the *IFNG* genic region, using the “ATG position” annotation (see Materials and Methods for details). Absolute frequencies, in numbers of chromosomes observed, of each haplotype lineage in Africa, Europe and East-Asia are reported. For each SNP, the first allele corresponds to the most parsimonious ancestral allele.

Supp. Table S1. Populations belonging to the HGDP-CEPH resequencing sub-panel

| Population | Geographical origin | Region | Number of individuals |
|-----------------------------------|---------------------|----------------------------------|-----------------------|
| Bantu | South Africa | sub-Saharan Africa | 8 |
| Bantu | Kenya | sub-Saharan Africa | 11 |
| Yoruba | Nigeria | sub-Saharan Africa | 22 |
| Mandenka | Senegal | sub-Saharan Africa | 21 |
| <i>sub-Saharan African</i> | | <i>sub-Saharan Africa</i> | <i>62</i> |
| Adygei | Russia Caucasus | Europe | 7 |
| Russian | Russia | Europe | 15 |
| French | France | Europe | 8 |
| French Basque | France | Europe | 12 |
| Orcadian | Orkney Islands | Europe | 6 |
| North Italian | Italy (Bergamo) | Europe | 3 |
| Sardinian | Italy | Europe | 11 |
| <i>European</i> | | <i>Europe</i> | <i>62</i> |
| Han | China | Asia | 15 |
| Dai | China | Asia | 2 |
| Lahu | China | Asia | 2 |
| Naxi | China | Asia | 3 |
| She | China | Asia | 3 |
| Yizu | China | Asia | 2 |
| Miaozu | China | Asia | 4 |
| Tujia | China | Asia | 1 |
| Tu | China | Asia | 2 |
| Xibo | China | Asia | 4 |
| Hezhen | China | Asia | 3 |
| Mongola | China | Asia | 4 |
| Daur | China | Asia | 2 |
| Oroqen | China | Asia | 1 |
| Cambodian | Cambodia | Asia | 4 |
| Japanese | Japan | Asia | 10 |
| <i>Asian</i> | | <i>Asia</i> | <i>62</i> |

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

Supp. Table S2. Details on resequenced regions and fragments for the *IFNG*, *IFNGR1* and *IFNGR2* genes.

| Gene | Chromosomal location | Sequenced fragments | Sequenced length | Exonic | Non-exonic |
|---|------------------------------|---------------------|------------------|--------|------------|
| <i>IFNG</i> NM_000619.2 NG_015840.1 | chr12:68,548,550-68,553,521 | -596 : 900 | 4,576 | 1,210 | 3,366 |
| | | 988 : 2,305 | | | |
| | | 3,586 : 5,347 | | | |
| <i>IFNGR1</i> NM_000416.2 NG_007394.1 | chr6:137,518,621-137,540,567 | -261 : 333 | 5,479 | 2,059 | 3,420 |
| | | 12,159 : 13,533 | | | |
| | | 14,726 : 16,051 | | | |
| | | 17,435 : 17,637 | | | |
| | | 17,898 : 18,639 | | | |
| 20,618 : 21,856 | | | | | |
| <i>IFNGR2</i> NM_005534.3 NG_007570.1 | chr21:34,775,202-34,809,827 | 7,146 : 7,709 | 4,777 | 1,555 | 3,222 |
| | | 10,950 : 11,667 | | | |
| | | 17,221 : 18,260 | | | |
| | | 22,990 : 23,613 | | | |
| | | 28,502 : 29,361 | | | |
| | | 33,038 : 33,192 | | | |
| 33,275 : 34,090 | | | | | |

Chromosome location is given according to the hg19 (GRCh37) human assembly coordinates. Positions are relative to the start coding site of the corresponding gene. Lengths are given in base pairs.

Supp. Table S3. Full list of polymorphisms found at *IFNG*, *IFNGR1* and *IFNGR2* genes.

| Gene | Chromosomal position | c.SNP | ATG position | Location | Protein domain | Amino-acid change | Polyphen | Panther | dbSNP | Hap MapII | Hap MapIII | AF | EU | AS | Singleton |
|-------------|-------------------------|-------------------|--------------|----------|----------------|-------------------|----------|---------|-------------|-----------|------------|------|------|------|-----------|
| <i>IFNG</i> | chr12:68553766-68553766 | c.-371T>A | -371 | 5' | | | | | rs3814242 | | | | | | 1.6 |
| <i>IFNG</i> | chr12:68553703-68553703 | c.-308G>T | -308 | 5' | | | | | rs2069709 | x | x | 5.6 | | | |
| <i>IFNG</i> | chr12:68552522-68552522 | c.115-483A>T | +874 | intron | | | | | rs2430561 | | | 18.5 | 54.0 | 17.7 | |
| <i>IFNG</i> | chr12:68552208-68552208 | c.115-169C>T | +1188 | intron | | | | | rs117641733 | | | | 0.8 | | Fre |
| <i>IFNG</i> | chr12:68552195-68552195 | c.115-156A>G | +1201 | intron | | | | | | | | | 0.8 | | Fre |
| <i>IFNG</i> | chr12:68551931-68551931 | c.183+40T>C | +1465 | intron | | | | | rs74099944 | | | 2.4 | | | |
| <i>IFNG</i> | chr12:68551554-68551554 | c.366+139T>C | +1842 | intron | | | | | | | | 0.8 | | | Ban |
| <i>IFNG</i> | chr12:68551409-68551409 | c.366+284G>A | +1987 | intron | | | | | rs1861494 | x | | 87.1 | 73.4 | 71.8 | |
| <i>IFNG</i> | chr12:68551343-68551343 | c.366+350A>T | +2053 | intron | | | | | | | | 0.8 | | | Ban |
| <i>IFNG</i> | chr12:68551333-68551333 | c.366+360G>C | +2063 | intron | | | | | | | | 0.8 | | | Ban |
| <i>IFNG</i> | chr12:68551196-68551196 | c.366+497T>C | +2200 | intron | | | | | rs1861493 | x | | 8.9 | 26.6 | 29.0 | |
| <i>IFNG</i> | chr12:68549786-68549786 | c.367-519A>G | +3610 | intron | | | | | rs2069719 | x | x | 97.6 | 100 | 100 | |
| <i>IFNG</i> | chr12:68549767-68549767 | c.367-500C>T | +3629 | intron | | | | | rs116174811 | | | 0.8 | | | Man |
| <i>IFNG</i> | chr12:68549710-68549710 | c.367-443G>A | +3686 | intron | | | | | rs2069720 | x | | 4.0 | | | |
| <i>IFNG</i> | chr12:68549686-68549686 | c.367-419T>C | +3710 | intron | | | | | | | | | | 0.8 | Oro |
| <i>IFNG</i> | chr12:68549549-68549549 | c.367-282T>A | +3847 | intron | | | | | | | | | 0.8 | | Sar |
| <i>IFNG</i> | chr12:68549494-68549494 | c.367-227G>C | +3902 | intron | | | | | | | | 0.8 | | | Man |
| <i>IFNG</i> | chr12:68549377-68549377 | c.367-110G>T | +4019 | intron | | | | | rs55662249 | | | | 1.6 | | |
| <i>IFNG</i> | chr12:68549033-68549036 | c.*97_*100delTCAA | +4360 | 3'UTR | | | | | | | | 1.6 | | | |
| <i>IFNG</i> | chr12:68548953-68548953 | c.*180C>T | +4443 | 3'UTR | | | | | rs2069722 | x | x | 5.6 | | | |
| <i>IFNG</i> | chr12:68548770-68548770 | c.*363G>A | +4626 | 3'UTR | | | | | rs55991209 | | | 1.6 | | | |
| <i>IFNG</i> | chr12:68548680-68548680 | c.*453G>C | +4716 | 3'UTR | | | | | rs7957366 | | | 0.8 | | | Ban |

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

| | | | | | | | | | | |
|---------------|-------------------------|-----------|--------|--------|-----------|---|------|------|------|--|
| <i>IFNGR2</i> | chr21:34809686-34809686 | c.*417G>A | +33837 | 3' UTR | rs12655 | | 3.2 | | | |
| <i>IFNGR2</i> | chr21:34809693-34809693 | c.*424T>C | +33844 | 3' UTR | rs1059293 | x | 12.9 | 55.6 | 15.3 | |

The position of each SNP was determined using the reference sequence listed in Supp. Table S2. The first amino-acid corresponds to the ancestral state, as defined considering the sequences of human, chimpanzee, gorilla, orangutan and rhesus. Chromosome location of each SNP is given according to the hg19 (GRCh37) human assembly. Protein domains are given by the UniProt Database. For the *IFNGR2* SNPs c.708A>T and c.889G>A, the domain is predicted by the UniProt Database as a potential domain. For the Panther analyses, the P-deleterious values are shown. The "x" means that the SNP is genotyped in HapMap Phase III and/or Phase II. The frequencies, given in %, of each SNP in the different merged continental populations refer to the Derived Allele Frequency in Africans (AF), Europeans (EU) and Asians (AS) of our study. Singletons observed are reported per individual population in our population sample; Ady: Adygei. Ban: Bantu. Dau: Daur. Cam: Cambodian. Fre: French. Hez: Hezhen. Jap: Japanese. Man: Mandenka. Nax: Naxi. Orc: Orcadian. Rus: Russian. Sar: Sardinian. Yor: Yoruba.

1
2
3 **Supp. Table S4: Genotype frequencies of each SNP in our population panel (HGDP-CEPH**
4 **sub-panel) as well as in the HapMap Phase II and III populations.** This table can be found as
5
6
7
8 Supp. Table S4.xls file.
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

